Confidence sets for change-point problems in nonparametric regression

Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultäten
vorgelegt dem
Fachbereich Mathematik und Informatik
der Philipps-Universität Marburg

von

M. Sc. Viktor Bengs

geboren in Kaptschagaj

Marburg, 2018

Vom Fachbereich Mathematik und Informatik
der Philipps-Universität Marburg (Hochschulkennziffer 1180)
als Dissertation angenommen am 18. September 2018

Erstgutachter: Prof. Dr. Hajo Holzmann
Zweitgutachter: Prof. Dr. Natalie Neumeyer

Tag der Einreichung: 12. Juli 2018
Tag der Disputation: 27. September 2018

In loving memory of my grandmother Lilia.

Acknowledgments

First and foremost, I want to thank my supervisor, Hajo Holzmann, for his superb support at any time, for the enlightening discussions we have had and for his honest criticism which improved this thesis as well as broadened my scientific knowledge significantly. Moreover, I thank Prof. Dr. Natalie Neumeyer for making the second assessment of this thesis. I thank my fellow working group members and my colleagues for the throughout nice working atmosphere, the interesting talks during lunchtime, the funny board game parties and especially for enduring my uncountable many bad jokes. Special thanks to Jasmin Hill and Heiko Werner for reading parts of this thesis and giving valuable comments as well as helpful corrections. Great thanks to my family and my friends for always bolstering me up. Thank you, Jasmin, for being there for me at any time. I gratefully acknowledge financial support from the DFG, grant Ho 3260/5-1.

Contents

Zusammenfassung

Nichtparametrische Regression hat über die letzten Jahre in der Statistik und anderen Bereichen große Aufmerksamkeit genossen. Die Anwendungsbereiche sind vielfältig und spielen für die Modellierung von Zusammenhängen zwischen verschiedenen Merkmalen eine große Rolle. Insbesondere wird oftmals mittels der nichtparametrischen Regression ein digitales Bild modelliert, indem die Bildintensitätsfunktion als Regressionsfunktion verwendet wird und die Pixel als erklärende Variablen. Dadurch besteht ein starker Zusammenhang zu dem Bereich der Bildverarbeitung.
Eine weiterer wichtiger Teilbereich von Regressionsproblemen sind solche, bei denen Irregularitäten in der Regressionsfunktion vorhanden sind, sodass die klassischen Glattheitsannahmen über den gesamten Definitionsbereich nicht gelten können. Solche Irregularitäten werden in eindimensionalen Regressionsproblemen change-points genannt und change-curves im multivariaten Fall. Sprünge oder Knicke in der Regressionsfunktion sind Beispiele für change-points, wohingegen change-curves in der Regel Unstetigkeitskurven entsprechen. Unstetigkeitskurven einer Bildintensitätsfunktion für zweidimensionale Bilder werden Kanten genannt und stellen ein allgegenwärtiges Objekt in diesem Bereich dar.
Klassische Glättungsverfahren wie zum Beispiel Kernregressions-, lokale Polynom- oder Waveletprojektionsschätzer sind für Regressionsprobleme mit change-points beziehungsweise change-curves nicht mehr optimal. Dies liegt daran, dass diese Methoden bei der Entfernung des Rauschens der Beobachtungen die change-points als solches Rauschen missinterpretieren würden und diese fälschlicherweise mitglätten. Dies würde in schlechten Schätzwerten für die Regressionsfunktion in der Nähe der change-points resultieren.
Um die Nachteile der klassischen Glättungsverfahren zu kompensieren, wurden in der Literatur alternative Methoden entwickelt. Insbesondere spielen für diese Methoden die explizite Schätzung der Stellen, der Ausprägung und der Anzahl der change-points eine zentrale Rolle, da diese wichtige Merkmale der Regressionsfunktion darstellen, die für eine adäquate Schätzung der Regressionsfunktion berücksichtigt werden müssen. Diese Arbeit befasst sich primär mit der Schätzung der Stellen, an denen change-points auftreten, sowie der damit verbundenen Unsicherheit der Schätzung, welche durch Konfidenzmengen erfasst wird. Entsprechend beschränkt sich die Diskussion der Literatur auf diese Themen.

Der mit Abstand meist studierte Typ von change-points ist eine Sprungstelle in der Funktion von Interesse. Eine weitverbreitete Methode um Sprungstellen in eindimensionalen Regressionsproblemen zu schätzen bzw. auf Sprungstellen zu testen, ist die geglättete Differenzenmethode, welche in Qiu und Li (1991), Müller (1992) bzw. Wu und Chu (1993) eingeführt wurde. Die Grundidee dabei ist, dass die absolute Differenz zwischen einem linksseitigen Kernschätzer und einem rechtsseitigen Kernschätzer in Bereichen, in denen die Funktion glatt ist, klein ist, jedoch in der Nähe von Sprüngen große Werte annimmt. Entsprechend ist die Stelle, an dem diese Differenz maximal wird, ein sinnvoller Schätzer für die Sprungstelle.
Dieser Ansatz ist für den eindimensionalen Fall gut geeignet, jedoch ist es nicht ohne weiteres möglich, dieses Konzept auf den mehrdimensionalen Fall zu erweitern. Dies liegt daran, dass im

mehrdimensionalen Fall überabzählbar viele Richtungen um eine Sprungkurve existieren, wohingegen es im eindimensionalen Fall nur zwei Richtungen auf der reellen Zahlengerade gibt. Um dieses Problem anzugehen, haben Qiu (1997) bzw. Müller und Song (1994) die rotierende geglättete Differenzenmethode entwickelt, welche es erlaubt, die Differenz der einseitigen Kerne in eine angemessene Richtung zu rotieren. Dadurch kann die Sprungkurve adäquat geschätzt werden, sodass eine sinnvolle Rekonstruktion der nicht-glatten Regressionsfunktion unter Berücksichtigung der geschätzten Sprungkurven möglich ist.

Abgesehen von der Rekonstruktion bzw. der Schätzung eines Objektes von Interesse ermöglicht statistische Modellierung die Konstruktion von Konfidenzmengen, welche das Objekt mit einer großen Wahrscheinlichkeit enthalten. Konfidenzmengen sind mittlerweile für verschiedene Probleme der nichtparametrischen Statistik gut erforscht, wie z.B. für nichtparametrische Dichteschätzung (Bickel und Rosenblatt (1973), Giné und Nickl (2010), Chernozhukov et al. (2014)), für glatte Regressionsfunktionen (Eubank und Speckman (1993), Neumann und Polzehl (1998), Proksch (2016)), und für inverse Probleme, sowie dem Messfehler-Modell (Bissantz und Holzmann (2008), Birke et al. (2010), Proksch et al. (2015), Delaigle et al. (2015)). Konfidenzbereiche für geometrische Merkmale wie Dichtelevelmengen (Mammen und Polonik (2013)) oder Moduskurven von Dichten (Qiao und Polonik (2016)) wurden ebenfalls bereits studiert. Konfidenzintervalle für Sprungstellen in eindimensionalen Regressionsproblemen wurden von Loader (1996), Gijbels et al. (2004), sowie Seijo und Sen (2011) konstruiert.
Allerdings gibt es bislang keine Methoden, um Konfidenzbereiche für Sprungkurven in bivariaten bzw. multivariaten Regressionsproblemen zu konstruieren. Um diese Lücke zu schließen, werden in Kapitel 2 gleichmäßige und punktweise asymptotische Konfidenzbänder für eine einzelne Sprungkurve in einer ansonsten glatten Regressionskurve konstruiert. Der Ausgangspunkt dafür ist die rotierende geglättete Differenzenmethode, so dass ein Bezug zu Qiu (2002) besteht, jedoch wird anstelle einer Schwellwertbildung eine Linearisierung einer lokalisierten Version des Kontrastes verwendet, um eine leichtere Anwendung der Techniken aus der M-Schätztheorie zu ermöglichen. Die Konstruktion der gleichmäßigen Konfidenzbänder beruht auf einer Gaußschen Approximation des Score-Prozesses und einer Anti-Konzentrationsungleichung für das Supremum von Gaußschen Prozessen ähnlich wie in Chernozhukov et al. (2014). Für die punktweisen Konfidenzbänder wird asymptotische Normalität der Schätzer verifiziert. Um die Anwendbarkeit der Ergebnisse auf endliche Datenmengen zu illustrieren, wird sowohl eine Simulationsstudie für künstliche Daten durchgeführt, als auch eine Anwendung der Methode auf ein digitales verrauschtes Bild.

Neben Sprungstellen in der Regressionsfunktion selbst gibt es weitere Irregularitäten, die in der sonst glatten Regressionsfunktion auftreten können. Ein weiterer relevanter Typ von Irregularitäten ist ein Sprung in einer der Ableitungen der Regressionsfunktion. Solch ein Punkt hat teilweise auch einen starken Einfluss auf die Form der Regressionsfunktion, da zum Beispiel eine Unstetigkeit in der ersten Ableitung zu einer abrupten Änderung der Steigung der Regressionsfunktion führt, oder eine Unstetigkeit in der zweiten Ableitung in einer plötzlichen Änderung der Krümmung resultiert. Wir nennen Sprünge in der γ-ten Ableitung γ-Knicke oder Knicke der Ordnung γ.
Die Schätzung von γ-Knicken war Gegenstand der Arbeiten von Müller (1992), Eubank und Speckman (1994), Wang (1995), Goldenshluger et al. (2006), Goldenshluger et al. (2008a), Cheng und Raimondo (2008), Wishart (2009), Wishart und Kulik (2010) als auch Mallik et al. (2013). Jedoch hat nur Mallik et al. (2013) explizit Konfidenzintervalle für die Stelle eines Knickes höherer Ordnung konstruiert. Die Annahmen an die Glattheit der Regressionsfunktion außerhalb des Knickes sind schwächer als in der Literatur zuvor, allerdings werden einschränkende Annahmen an die Form der Regressionsfunktion gemacht, welche die Anwendbarkeit ihrer Methode limitiert.
Eine effiziente Technik zur Schätzung der Knick-Stelle ist die sogenannte *zero-crossing-time-technique*, welche die Nullstelle einer geglätteten zweiten Ableitung der Funktion von Interesse als Schätzer benutzt. Die Grundidee ist, dass falls die Funktion von Interesse einen Sprung in θ auf-

weist, so hat eine geeignete geglättete Version dieser Funktion eine große Steigung in der Nähe von θ. Entsprechend hat die erste Ableitung dieser geglätteten Version ein lokales Maximum in der Nähe von θ und die zweite Ableitung dort eine Nullstelle. Die Methode besteht dann aus zwei Phasen: In der ersten Phase wird ein Intervall bestimmt, welches den Knick mit einer hohen Wahrscheinlichkeit enthält. In der zweiten Phase wird der Knick durch eine Nullstelle einer empirischen Version der geglätteten zweiten Ableitung innerhalb des Intervalls geschätzt. Goldenshluger et al. (2006) haben diese Methode verwendet, um die einzelne Sprungstelle in einem indirekten Modell in weißem Rauschen zu schätzen. Zusätzlich haben sie die Minimax-Optimalität dieser Methode verifiziert, falls die Signalfunktion mindestens Lipschitz-glatt außerhalb der Sprungstelle ist. Um diese Methode auch für ein direktes Problem anwenden zu können, haben Cheng und Raimondo (2008) explizite Kerne für die Glättung der zweiten Ableitung konstruiert, um auf Knicke in der Regressionsfunktion zu testen. Wishart (2009) bzw. Wishart und Kulik (2010) verwendeten dann letztere Modifikation um Knicke zu schätzen, falls das Design oder die Fehler im Regressionsproblem Abhängigkeitsstrukturen aufweisen. Zwar wurde die Optimalität der Schätzmethode in den gerade genannten Papieren rigoros studiert, jedoch wurden asymptotische Verteilungseigenschaften gänzlich außer Acht gelassen. Dies liegt an der verwendeten Charakterisierung des Schätzers als M-Schätzer, anstatt die natürlich gegebene Charakterisierung als Z-Schätzer zu nutzen, so dass Konvergenzraten als M-Schätzer leicht hergeleitet werden können, aber asymptotische Verteilungsergebnisse eher schwer nachweisbar sind. In Kapitel 3 wird die Charakterisierung als Z-Schätzer verwendet, um die asymptotische Normalität des Schätzers herzuleiten und somit Konfidenzmengen zu konstruieren. Basierend auf der asymptotischen Normalität und einer Lepski-Wahl des Tuningparameters der zero-crossing-time-technique werden für einen einzelnen γ-Knick adaptive Konfidenzintervalle konstruiert, die sich an die Hölder-Glattheit der Regressionsfunktion außerhalb des Knickes anpassen. Eine Simulationsstudie für künstliche Daten und einige Anwendungen auf reale Daten unterstreichen die Nützlichkeit der zuvor hergeleiteten Ergebnisse.

Diese Arbeit ist wie folgt strukturiert. Das erste Kapitel gibt eine prägnante Einführung in die nötigen theoretischen Konzepte für diese Arbeit. Zudem wird der aktuelle Forschungsstand in der Literatur bezüglich der Schätzmethoden und der Konstruktion von Konfidenzmengen für change-points diskutiert. Kapitel 2 konstruiert gleichmäßige und punktweise Konfidenzbänder für eine einzelne Sprungkurve in einem bivariaten Regressionsproblem mittels M-Schätzmethoden und Gaußscher Approximation. Die Resultate basieren größtenteils auf Bengs et al. (2018). Adaptive Konfidenzintervalle für einen einzelnen Knick höherer Ordnung werden in Kapitel 3 behandelt, welches auf Bengs und Holzmann (2018) beruht. Kapitel 4 widmet sich der Herleitung der Minimax-Konvergenzraten für die Modelle in den beiden Kapiteln zuvor, um deren Optimalität zu zeigen. Die optimale Konvergenzrate für das Modell in Kapitel 2 ist bereits bekannt, wohingegen die für das Modell in Kapitel 3 neu ist. Schließlich werden im Anhang verschiedene Hilfsresultate bereitgestellt, welche von unabhängigem Interesse sein können, da diese zum einen bequeme Hilfsresultate für nichtparametrische Regressionsprobleme mit deterministischem Design beinhalten und zum anderen Ergebnisse der klassischen Wahrscheinlichkeitstheorie auf ein Szenario, in dem gleichmäßige Ergebnisse von Interesse sind, erweitern.

Introduction

Nonparametric regression has been a topic of major interest over the last years in statistics and other fields. Its application is manifold and plays nowadays an important role for modeling relationships between different features as well as for denoising. In particular, nonparametric regression is often used for modeling an image by interpreting the regression function as the corresponding intensity function of the image and the pixels as the explaining variables. Consequently, there is a strong relationship to the realm of image processing.
An important subdomain of regression problems are frameworks where regression functions with irregularities emerge, such that global smoothness assumptions (in some specific sense) as required in the classical nonparametric regression cannot hold over the whole domain space. Such irregularities are called change-points in univariate settings or change-curves in the multivariate extension. Change-points can be for instance jumps or kinks of the regression function, while change-curves usually correspond to discontinuity curves. In two-dimensional image functions, a discontinuity curve of the intensity function of the image is called an edge and is an omnipresent object in this area.
For regression problems with change-points or change-curves the classical smoothing procedures such as kernel-smoothing, local-polynomial or wavelet projection to name a few, are often no longer optimal in the presence of such irregularities. In order to denoise the observations the change-point-locations would be misconceived as noise and consequently be smoothed out at these locations. The resulting estimators of the regression function would not give a good fit near these change-point-locations.
Alternative procedures have been proposed to circumvent the drawbacks of classical smoothing procedures in a nonparametric regression framework with change-points. In particular, the explicit estimation of the location, the magnitude as well as the number of such change-points play a major role in this realm, as these are specific characteristics of the regression function, which need to be incorporated in order to guarantee a suitable estimation of the non-smooth regression function. The main focus of this thesis is on estimation of change-point-locations and the corresponding uncertainty of this estimation problem expressed through confidence sets, so that discussion of the literature concentrates mainly on this particular issue.

The most studied type of change-point is a jump-discontinuity of a function of interest. One popular method to detect or to estimate jump-point-locations in univariate frameworks is the so called smoothed difference approach which was introduced in Qiu and Li (1991), Müller (1992) resp. Wu and Chu (1993). The basic idea is that the absolute difference of a left-sided kernel estimator and right-sided kernel estimator should be nearly zero at points where the function is smooth, but should be large near a jump-point. Consequently, a sensible estimator for the jump-point-location is given by the point maximizing this difference.
Although this procedure works well in the one-dimensional case, it is not straightforward to extend this method to the bivariate case. This is due to the uncountable infinite many directions around a jump-location-curve in the bivariate setting, while in a univariate setting there are only two directions around the jump-point on the real line. To this end, Qiu (1997) or Müller and Song (1994) proposed

the rotated difference kernel estimation which allows to rotate the difference of the one-sided kernel estimators in an appropriate direction. This extension allows a proper estimation of the jump-location-curve and in addition a sensible reconstruction of the regression function by incorporating these locations.

Apart from mere reconstruction or estimation, statistical modeling allows for the construction of confidence sets, in which the object of interest is located with high probability. Confidence sets are by now well-developed for various problems in nonparametric statistics, e.g. for nonparametric density estimation (Bickel and Rosenblatt, 1973; Giné and Nickl, 2010; Chernozhukov et al., 2014), for smooth regression functions (Eubank and Speckman, 1993; Neumann and Polzehl, 1998), and for deconvolution and errors-in-variables problems (Bissantz and Holzmann, 2008; Birke et al., 2010; Proksch et al., 2015; Delaigle et al., 2015). Mammen and Polonik (2013) and Qiao and Polonik (2016) focus on more geometrical features, and construct confidence regions for density level sets and the density ridge, respectively. Confidence intervals for jump-points in univariate regression settings were constructed by Loader (1996), Gijbels et al. (2004) as well as Seijo and Sen (2011).

However, currently there seem to be no methods available to construct a confidence set for the discontinuity curve of a regression function in the bivariate case. Therefore, Chapter 2 is devoted to the construction of uniform and pointwise asymptotic confidence bands for the single edge in an otherwise smooth image function based on the rotational difference kernel estimator, and hence is also related to Qiu (2002), but instead of thresholding, a linearization of a localized version of this contrast is developed to use the convenience of M-estimation methods. The uniform confidence bands then rely on a Gaussian approximation of the score process together with anti-concentration results for suprema of Gaussian processes from Chernozhukov et al. (2014), while pointwise bands are based on asymptotic normality. A simulation study for investigation of the finite-sample performance of the proposed methods is provided as well as an illustrative application of the proposed method to a real-world image.

Besides jump discontinuities in the regression function itself there are other types of irregularities that might occur in an otherwise smooth regression function. Another certainly relevant type of irregularity is a jump-point in some derivative of the regression function. Such a point can have an extraordinary impact on the shape of the regression function, e.g. a jump discontinuity in the first derivative would relate to an abrupt change in the direction of the regression function, while a jump discontinuity in the second derivative describes a sudden change in the curvature. We refer to a jump in the γ-th derivative as a γ-kink or a kink of order γ.

Estimation of γ-kink-locations were considered by Müller (1992), Eubank and Speckman (1994), Wang (1995), Goldenshluger et al. (2006), Goldenshluger et al. (2008a), Cheng and Raimondo (2008), Wishart (2009), Wishart and Kulik (2010) resp. Mallik et al. (2013). However, only Mallik et al. (2013) dealt explicitly with the construction of confidence intervals for the location of a single kink of higher order. Their assumptions on the smoothness of the regression function are milder than those made in the aforementioned literature on kink estimation. Though they require some shape conditions on the regression function which restricts the applicability of their method.

An efficient method to estimate kink-locations is the so-called zero-crossing-time-technique, which uses the zero of a smoothed second derivative as an estimate. The main idea is that if the function of interest has a jump-location, say at θ, a smoothed version of the function will have a large slope near θ, so that its first and second derivatives have a local maximum respectively a zero near θ. Additionally, the zero-crossing-time-technique consists of two stages: In the first stage an interval which entails the kink-location with high probability is constructed, while the second stage estimates the kink-location by a zero of the empirical smoothed second derivative inside this interval. Goldenshluger et al. (2006) used the zero-crossing-time-technique for estimating a single jump-point-location in an indirect white noise model and showed its minimax optimality only assuming that the function of interest is at least Lipschitz continuous away from the jump-point-location. The adaptation of this method to a direct setting was made by Cheng and Raimondo (2008) to detect kink-locations, as well

as by Wishart (2009) resp. Wishart and Kulik (2010) to incorporate dependency structures in the errors resp. the design.
Although the optimality of this estimate is thoroughly investigated in these papers, none of them studies the asymptotic distribution of this estimate. This is due to characterization of their estimate as an M-estimate instead of the naturally characterization as a Z-estimate, which is along their lines suitable to obtain rates of convergence, but intricate to show for instance asymptotic normality.
In Chapter 3 the characterization as a Z-estimate is used to show asymptotic normality of the estimate and consequently to construct confidence sets by modifying the involved kernel smoothers. Moreover, based on the asymptotic normality and on a Lepski-choice of the resolution level in the zero-crossing-time-technique, adaptive confidence intervals are constructed for a single kink of order γ with respect to the Hölder-smoothness of the regression function away from the kink. Simulation studies for artificially constructed data and for common real-world datasets in this realm are provided as well.

This thesis is structured as follows. The first chapter introduces concisely the involved theoretical concepts for the remainder of this thesis and the exposition is in large parts of tutorial nature. Moreover, Chapter 1 provides a review on the literature about the state-of-the-art of estimation procedures and construction of confidence sets for change-point-locations in nonparametric regression settings. Chapter 2 develops uniform and pointwise asymptotic confidence bands for the jump-location-curve in a boundary fragment model using methods from M-estimation and Gaussian approximation. These results are based on Bengs et al. (2018). Construction of adaptive confidence intervals for the single kink-location of higher order is covered in Chapter 3, which is based on Bengs and Holzmann (2018). Chapter 4 is devoted to derive the minimax rate of convergence for the models in the chapters before and deducing their optimality. The minimax-optimal rate of convergence for the model in Chapter 2 is well-known, while the optimal rate of convergence for the model in Chapter 3 is new. Finally, the appendix provides various results which could be of independent interest, especially for analyzing nonparametric regression problems with fixed design, as well as for extending the classical probability theory results to a uniform framework.

Notation

We shall use the following notation. We write P_θ resp. $\mathbb{E}_\theta$ to denote the probability measure resp. the expected value with respect to some parameter θ. If a sequence of random variables $X_1, X_2, \ldots$ converges in probability to a random variable X we write $X_n \xrightarrow{\mathrm{P}} X$ or $X_n = X + o_P(1)$. If the convergence is in distribution we write $X_n \rightsquigarrow X$. In addition, for random variables X, Y we mean by $X \stackrel{d}{=} Y$ equality in distribution. We say that some event $A = A(n)$ occurs *with high probability* if $P(A(n)) \to 1$ provided $n \to \infty$. We say that some event $A = A(n)$ *occurs with high probability and uniformly over some set* $\mathcal{F}$ if $\inf_{f \in \mathcal{F}} P_f(A(n)) \to 1$.
Let λ_d denote the d-dimensional Lebesgue-measure. For $d \in \mathbb{N}$, $\mu \in \mathbb{R}^d, \Sigma \in \mathbb{R}^{d \times d}$ symmetric and positive definite, let $N_d(\mu, \Sigma)$ be the normal distribution with expectation μ and covariance matrix Σ. For a vector $a \in \mathbb{R}^d$ we denote by $\operatorname{diag}(a)$ the $d \times d$ diagonal matrix with diagonal entries a, while I_d is the $d \times d$ identity matrix. Φ denotes the cumulative distribution of the standard normal distribution. Given $\alpha \in (0,1)$ we denote by $q_\alpha(X)$ the α-quantile of the distribution of a random variable X resp. by $q_\alpha(Q)$ the α-quantile of a distribution Q. For measures μ and ν on some measurable space $(\mathcal{X}, \mathcal{A})$ we write $\mu \ll \nu$ if μ is absolutely continuous with respect to ν and the corresponding Radon-Nikodym-derivative is $d\mu/d\nu$.

The function $\Pi_i : \mathbb{R}^d \to \mathbb{R}$ is the projection onto the i-th coordinate and in particular for a vector $\mathbf{z} = (z_1, \ldots, z_d)^T \in \mathbb{R}^d$ we denote the coordinate projection onto the i-th coordinate as $(\mathbf{z})_i = \Pi_i(\mathbf{z}) = z_i$ for $i = 1, \ldots, d$. Furthermore, we write $\mathbf{z}^\alpha = z_1^{\alpha_1} \ldots z_d^{\alpha_d}$ for $\alpha = (\alpha_1, \ldots, \alpha_d)^T \in \mathbb{N}^d$ and

$$f^{(\alpha)}(\mathbf{z}) = \frac{\partial^{\alpha_1 + \ldots + \alpha_d} f}{\partial z_1^{\alpha_1} \ldots \partial z_d^{\alpha_d}}(\mathbf{z})$$

for a two times continuously differentiable function $f : \mathbb{R}^d \to \mathbb{R}$. In particular,

$$\nabla f(\mathbf{z}) = (f^{(1,0,\ldots,0)}, \ldots, f^{(0,\ldots,0,1)})^T(\mathbf{z})$$

and

$$\nabla \nabla^T f(\mathbf{z}) = \begin{pmatrix} f^{(2,0,\ldots,0)} & f^{(1,1,0,\ldots,0)} & \cdots & f^{(1,0,\ldots,0,1)} \\ f^{(1,1,0,\ldots,0)} & f^{(0,2,0\ldots,0)} & & \vdots \\ \vdots & & \ddots & \vdots \\ f^{(1,0,\ldots,0,1)} & & \cdots & f^{(0,\ldots,0,2)} \end{pmatrix}(\mathbf{z}).$$

We also write $\partial_{z_i} f(\mathbf{z})$ for $f^{\mathbf{e_i}^T}(\mathbf{z})$, where $\mathbf{e_i}$ is the i-th canonical unit vector. If $g : X \to Y$ for $X \subset \mathbb{R}^d$ and $Y \subset \mathbb{R}$ then we let $\operatorname{epi}(g) = \{(\mathbf{x}, y)^T \in X \times Y \mid g(\mathbf{x}) \leq y\}$ be the epigraph of g.

Let $A \triangle B$ denote the symmetric difference between two sets A and B, e.g.

$$A \triangle B := (A \backslash B) \cup (B \backslash A) = (A \cup B) \backslash (A \cap B), \quad A, B \subset \mathbb{R}^d.$$

Write $\mathbf{z}+A := \{\mathbf{z}+\mathbf{y} : \mathbf{y} \in A\}$ for $\mathbf{z} \in \mathbb{R}^d$ and a set $A \subset \mathbb{R}^d$. For real-valued sequences $(a_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ and $(b_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ we write $a_n \cong b_n$ if there exist finite constants $C_1, C_2 > 0$ and an $n_0 \in \mathbb{N}$ such that $C_1 \le |a_n/b_n| \le C_2$ for all $n \ge n_0$.

Let $\langle\cdot,\cdot\rangle$ be the Euclidean scalar product on $\mathbb{R}^d$ and $||\cdot||_2$ the corresponding Euclidean-norm. Denote by $||\cdot||$ a norm on $\mathbb{R}^d$ as well as on $\mathbb{R}^{d\times d}$, where the dimension should be clear from the context. We only assume that the matrix-norm is compatible with the vector norm, that is $||\mathbf{Az}|| \le ||\mathbf{A}||\cdot||\mathbf{z}||$ for a matrix $\mathbf{A}$ and a vector $\mathbf{z}$ and that the matrix-norm is submultiplicative, i.e. $||\mathbf{AB}|| \le ||\mathbf{A}||\cdot||\mathbf{B}||$ for matrices $\mathbf{A},\mathbf{B}$. With a slight abuse of the notation we also denote the L_2-norm with $||\cdot||_2$. For a function $f : I \longrightarrow \mathbb{R}^{i\times j}$ for $i,j \in \{1,\ldots,d\}$ with either $i=j$ or $i \ge 1$ and $j=1$, and I is a compact subinterval of $\mathbb{R}^d$ we define the sup-norm as

$$||f||_\infty = \begin{cases} \sup_{x\in I} \ |f(x)|, & i=j=1, \\ \sup_{x\in I} ||f(x)||, & \text{else.} \end{cases}$$

Uniform Landau symbols

To express the uniformity of some results in this thesis in a concise way, we introduce an extension of the classical Landau notation in the following sense. Let $\mathcal{F}$ be some set, $T \subset \mathbb{R}$, $A \subset (0,\infty)$ and $g : (0,\infty) \longrightarrow (0,\infty)$. For a family of functions $(F_h)_{h>0}$ with $F_h : T\times\mathcal{F} \longrightarrow \mathbb{R}$, we write $F_h(t,f) = O_{f\in\mathcal{F},t\in T}(g(h))$ for $h \in A$ if and only if there exists an $M > 0$ such that for all $f \in \mathcal{F}$, $t \in T$ and any $h \in A$ holds $|F_h(t,f)| \le Mg(h)$. If the functions F_h are constant in t we just write $O_{\mathcal{F}}(g(h))$ for $O_{f\in\mathcal{F},t\in T}(g(h))$. Furthermore, we write $F_h(t,f) = o_{\mathcal{F}}(g(h))$ if and only if for all $\delta > 0$ there exists an $h_0 > 0$ such that for any $f \in \mathcal{F}$, $t \in T$ and any $h \in (0,h_0)$ holds $|F_h(t,f)| \le \delta g(h)$.

Next, let $\tilde{g} : \mathbb{N}\times(0,\infty) \longrightarrow (0,\infty)$ and now $A \subset \mathbb{N}\times(0,\infty)$. For a family of functions $(F_{n,h})_{n\in\mathbb{N},\,h>0}$ with $F_{n,h} : T\times\mathcal{F} \longrightarrow \mathbb{R}$, we write $F_{n,h}(t,f) = O_{f\in\mathcal{F},t\in T}(\tilde{g}(n,h))$ for $(n,h) \in A$ if and only if there exists an $M > 0$ such that for all $f \in \mathcal{F}$, $t \in T$ and any $(n,h) \in A$ holds $|F_{n,h}(t,f)| \le M\tilde{g}(n,h)$. If the functions $F_{n,h}$ are constant in t we just write $O_{\mathcal{F}}(\tilde{g}(n,h))$ for $O_{f\in\mathcal{F},t\in T}(\tilde{g}(n,h))$. In the same spirit as above we define $o_{\mathcal{F}}$ for this case.

Finally, for the stochastic counterparts, we have the following definitions if $\mathcal{F}$ is the parameter space. Then for a family of random functions $(\hat{F}_{n,h})_{n\in\mathbb{N},\,h>0}$ with $\hat{F}_{n,h} : \Omega\times T \longrightarrow \mathbb{R}$ we write $\hat{F}_{n,h}(t) = O_{P,\mathcal{F}}(\tilde{g}(n,h))$ for $(n,h) \in A$ if and only if for all $\delta > 0$ there exists a finite constant $M > 0$ such that

$$\sup_{f\in\mathcal{F}} P_f\left(|{}^{\hat{F}_{n,h}(t)}/_{\tilde{g}(n,h)}| > M\right) \le \delta, \quad \forall (n,h) \in A,$$

while we write $\hat{F}_{n,h}(t) = o_{P,\mathcal{F}}(\tilde{g}(n,h))$ if and only if for all $\delta > 0$ and all $\tilde{\delta} > 0$ there exists a pair $(n_0,h_0) \in \mathbb{N}\times(0,\infty)$ such that

$$\sup_{f\in\mathcal{F}} P_f\left(|{}^{\hat{F}_{n,h}(t)}/_{\tilde{g}(n,h)}| > \tilde{\delta}\right) \le \delta, \quad \forall n \ge n_0,\, h \in (0,h_0).$$

CHAPTER 1

Nonparametric regression and change-point problems

This thesis deals with the construction of confidence regions in nonparametric regression problems with change-points. In order to align the issue of this thesis, this chapter gives an overview of the relevant areas of statistics for our purposes. In particular, in what follows there will be a general motivation for the nonparametric regression model as well as an introduction to the relevant terms and notions in this setting and its relation to image analysis. To validate the performance of estimation methods the concept of minimax risk is described. Afterwards, change-point problems are introduced with a thorough review on the literature for change-point analysis under nonparametric regression with a special focus on change-point-location estimation. Following this, M- and Z-estimation are introduced in a more general setting than in the classical theory as in van der Vaart (2000), which will be fundamental for the further chapters of this thesis. Then, there is an extensive section about the construction and relevance of confidence sets in statistics. Finally, the state-of-art of constructing confidence sets for change-points in nonparametric regression problems is discussed to emphasize the contribution of this thesis to the current status of research.

1.1 Nonparametric regression and minimax estimation theory

Given some random variable $Y \in \mathbb{R}$ and some random variable $X \in D \subset \mathbb{R}^d$, one may ask if there is a relationship between Y and X and if this is the case, how this relationship can be modeled or described appropriately in a mathematical sense. The classical probabilistic approach for this purpose is to consider $m(x) = \mathbb{E}(Y|X = x)$ which gives rise to the model given by

$$Y = m(X) + \varepsilon. \tag{1.1}$$

In this context $m : D \longrightarrow \mathbb{R}$ is the so-called *regression function*, which is unknown and describes the relationship between Y and X up to some *error* in the shape of the random variable ε. The model in (1.1) implies that the error has zero mean which means (1.1) is observed without any error on average. The *regression problem* consists of statistical issues such as estimation of the regression function arising from observing samples $(X_{i_1,\ldots,i_d}, Y_{i_1,\ldots,i_d})$, $i_1, \ldots, i_d = 1, \ldots, n$ of (X, Y) based on (1.1), that is

$$Y_{i_1,\ldots,i_d} = m(X_{i_1,\ldots,i_d}) + \varepsilon_{i_1,\ldots,i_d}, \quad i_1, \ldots, i_d = 1, \ldots, n, \tag{1.2}$$

where $\varepsilon_{i_1,\ldots,i_d}$ are error variables with $\mathbb{E}(\varepsilon_{i_1,\ldots,i_d}|X_{i_1,\ldots,i_d}) = 0$. The model (1.2) is called *nonparametric regression problem (NPP)* if it is assumed that the regression function m belongs to an infinite dimensional parameter set $\mathcal{F}$, for instance the set of all Lipschitz-continuous functions. We refer to the random variables $X_{i_1,\ldots,i_d}$ as *design points* and the *design* is given by $\mathbf{X} = (X_{i_1,\ldots,i_d})_{i_1,\ldots,i_d=1,\ldots,n}$,

the collection of these design points. If the design points $X_{i_1,\dots,i_d}$ are deterministic, the design is called *deterministic* and it is custom to write $\mathbf{x}_{i_1,\dots,i_d}$ instead of $X_{i_1,\dots,i_d}$. Otherwise, if the design points are random variables the design is called *random*.
In this thesis we consider only the case of a deterministic design. Moreover, it will be assumed that the domain D of the regression function is some compact set in $\mathbb{R}^d$ and the design points form an equidistant grid on D, also referred to as an *equidistant design.*

Image representation

The bivariate version of model (1.2) with equidistant design is of special interest for image analysis, as it is convenient to describe an image as follows. The design points $\mathbf{x_{i_1,i_2}}$ correspond to the pixels of the image and $m(\mathbf{x_{i_1,i_2}})$ is the corresponding gray-levels of the image at the pixel $\mathbf{x_{i_1,i_2}}$. As digitized images often contain noise the true gray-levels cannot be observed directly, but only a noised version. Such noise occurs for instance by the image acquisition and digitization of the image to the storage. In this setting and for multivariate versions of model (1.2) with equidistant design it is common to refer to m as the *image function.* Allowing m to map into some multivariate space gives rise to modeling digital color images as well. Moreover, the dimension $d = 2$ is certainly the most prevalent for image analysis, whereas the case $d = 3$ is also of interest, as three-dimensional pictures are highly relevant for instance in medical applications.

Minimax optimality of an estimate

We do not restrict ourselves to estimation of the regression function only, but also consider estimation of other objects of interest which are related to the regression function. For instance, suppose we can decompose the regression function into finite many functions $f_1,\dots,f_m$, that is $m = f_m \circ \dots \circ f_1$ and we are interested in estimation of f_i for some $i = 1,\dots,m$. Consequently, we keep the following concise review on minimax optimality theory as general as possible.

Risk-functions

It is essential for statistical considerations to have a quantity which captures the reliability of an estimate. For this purpose the so-called *risk-functions* are introduced. Let $\mathcal{F}$ be some parameter space, possibly infinite dimensional, such that there exists a semi-distance d on $\mathcal{F}$. More precisely, $d : \mathcal{F} \times \mathcal{F} \to [0,\infty)$ satisfies for any $f, f', f'' \in \mathcal{F}$

1. $d(f,f) = 0$,
2. $d(f,f') = d(f',f)$,
3. $d(f,f'') \leq d(f,f') + d(f',f'')$.

Further, let $w : \mathbb{R}_+ \to \mathbb{R}_+$ be a *loss function,* that is $w(0) = 0$, w is monotone increasing and w is not the zero function. The *d-risk of an estimate* $\hat{f}_n \in \mathcal{F}$ for some specific $f \in \mathcal{F}$ is then defined by

$$\mathbb{E}_f\big(w(r_n^{-1} d(\hat{f}_n, f))\big),$$

where $r_n \subset \mathbb{R}^+$ is some sequence normalizing $d(\hat{f}_n, f)$, referred to as the *rate of convergence of* $\hat{f}_n$. Popular risks, which are of importance for this thesis, are the following

1. the *Mean-Integrated-Squared-Error* (MISE) given by $\mathbb{E}_f(r_n^{-2} \,||\, \hat{f}_n - f \,||_{L_2}^2)$ by choosing d as the L_2-metric and $w(x) = x^2$;
2. the *sup-norm-risk* defined as $\mathbb{E}_f(r_n^{-1} \,||\, \hat{f}_n - f \,||_\infty)$ by setting d as the L_∞-metric and $w \equiv 1$;

3. the *probability-risk* given by $P_f(r_n^{-1}d(\hat{f}_n,f) \geq A)$ by using $w(x) = 1_{x\geq A}$ for some $A > 0$.

Such d-risks are useful tools for the comparison of different estimators. However, the d-risk alone is not a satisfactory measure for the general statistical performance of some estimate $\hat{f}_n$, since it takes not into account how well different objects of interest, say $f_1,\ldots,f_m$, are estimated. Hence, it is more sensible to consider the *maximum d-risk* of an estimate $\hat{f}_n$ over $\mathcal{F}$, given by

$$R_w(\hat{f}_n;\mathcal{F},d,r_n) = \sup_{f\in\mathcal{F}} \mathbb{E}_f\left(w(r_n^{-1}d(\hat{f}_n,f))\right).$$

This quantity is mainly driven by the estimation properties of $\hat{f}_n$ on the parameter space $(\mathcal{F},d)$, which is in the realm of nonparametric statistics usually an infinite-dimensional parameter space such as a class of functions. More precisely, the typical choice for the parameter space is of the form

$$\mathcal{F} = \{f : I \to \mathbb{R} \,|\, f \text{ satisfies some smoothness conditions}\},$$

where $I \subset \mathbb{R}^d$ is some compact set. Some common function classes, which will be of interest for this thesis, are the following. Let $s \in \mathbb{N}$, then

$$C^s(I) = \{f : I \to \mathbb{R} \,|\, f \text{ is } s\text{-times continuous differentiable}\} \tag{1.3}$$

is the *function class of s-times on I continuous differentiable functions.* Given $s > 0$ we let $\lfloor s \rfloor = \max\{k \in \mathbb{N}_0 : \ k < s\}$, and we define the *Hölder class of functions on I* with smoothness parameter $s > 0$ and Hölder-constant $L > 0$ by

$$\mathcal{H}^s(I,L) = \left\{f \in C^{(\lfloor s \rfloor)}(I) \,|\, |f^{(\lfloor s \rfloor)}(\mathbf{x}) - f^{(\lfloor s \rfloor)}(\mathbf{y})| \leq L\,||\mathbf{x}-\mathbf{y}||_\infty^{s-\lfloor s \rfloor},\ \mathbf{x},\mathbf{y} \in I\right\}. \tag{1.4}$$

In particular, the *class of Lipschitz continuous functions on I* with Lipschitz constant $L > 0$ is

$$\mathrm{Lip}(I,L) = \mathcal{H}^1(I,L) = \left\{f \in C(I) \,|\, |f(\mathbf{x}) - f(\mathbf{y})| \leq L||\mathbf{x}-\mathbf{y}||_\infty,\ \mathbf{x},\mathbf{y} \in I\right\}. \tag{1.5}$$

In the cases (1.3) and (1.4) the quantity s measures the smoothness of the functions inside the class, while the additional term L is some regularity constant.

Minimax optimality

After verifying a rate of convergence r_n for a specific estimation procedure over some parameter space $(\mathcal{F},d)$, one wishes to know if there are estimation procedures which can attain a faster rate of convergence for this statistical model. For this purpose, one investigates the *minimax d-risk* for some parameter space $\mathcal{F}$ given by

$$\inf_{T_n} R_w(T_n;\mathcal{F},d,r_n) = \inf_{T_n} \sup_{f\in\mathcal{F}} \mathbb{E}_f\left(w(r_n^{-1}d(T_n,f))\right), \tag{1.6}$$

where the infimum is taken over all possible estimators T_n. Some sequence $(r_n)_n \subset \mathbb{R}^+$ is called *(minimax) rate* or *optimal rate of convergence* for estimators on $(\mathcal{F},d)$ if there exist finite constants $C_1 \geq C_2 > 0$ such that

$$\limsup_n \inf_{T_n} R_w(T_n;\mathcal{F},d,r_n) \leq C_1, \quad \text{and} \quad \liminf_n \inf_{T_n} r_n R_w(T_n;\mathcal{F},d,r_n) \geq C_2, \tag{1.7}$$

where the infimum is taken over all possible estimators T_n. An estimator $\hat{f}_n$ is called *rate optimal estimator* on $(\mathcal{F},d)$ if $R_w(\hat{f}_n;\mathcal{F},d,r_n) \leq C$ where $C > 0$ is some finite constant and r_n is the minimax rate for $(\mathcal{F},d)$. Section 4.1 provides a well-known approach for deriving optimal rates of convergence.

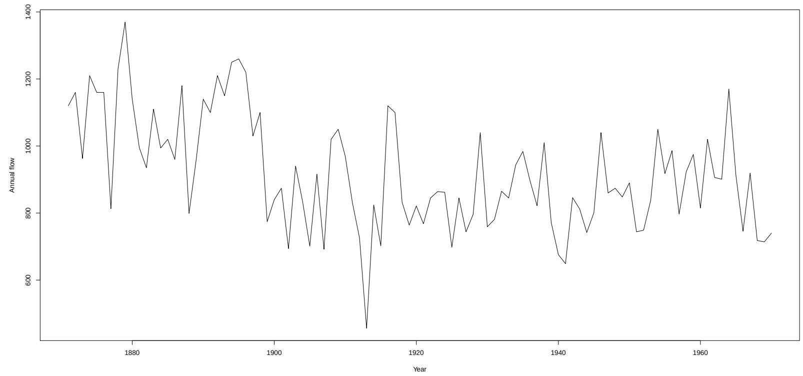

Figure 1.1.: Annual flow (in $10^8 m^3$) of the river Nile at Aswan from 1871 to 1970.

1.2 Change-point analysis under nonparametric regression

A large body of papers consider the nonparametric regression problem in (1.2) and assume that the regression function m is element of some function class containing only global smooth functions. Typical examples are the function classes in (1.3) – (1.5). It is often justifiable to assume that such a global smoothness of the regression function cannot hold for specific datasets. Consider Figure 1.1 for instance, where the annual flow of the Nile is displayed for the time from 1871 to 1970. Cobb (1978) pointed out that there is an apparent change-point visible around the year 1898, so that it is not appropriate to model the data with a regression function which is smooth over the whole domain. In addition, classical estimation procedures such as kernel smoothing, local linear fitting or wavelet estimation are not longer optimal in the presence of change-points without essential amendments.

Interpretation of a change-point

The first work in a regression setting with discontinuities interpreted as change-points is due to Thistlethwaite and Campbell (1960). Ever since change-points have attracted much attention in the research and many scientific articles from diverse academic disciplines dealt with problems of a change-point nature. As a consequence there have evolved different possibilities to declare what a change-point is. Casually speaking, a change-point of a function f is any point where f changes its local behavior suddenly and drastically. The most prevalent definition of a change-point is declaring it as a point of discontinuity in the function of interest, though this is quite not the only option. Discontinuities in the derivatives of the function of interest are as well points with potential strong impacts on the behavior of the function, e.g. a jump discontinuity in the first derivative would relate to an abrupt change in the direction of the regression function, while a jump discontinuity in the second derivative describes a sudden change in the curvature. Such an irregularity will be called a γ-*kink* or a kink of order γ, that is a jump in the γ-th derivative of the regression function. Generalizations of these kinks in a Hölder sense are γ-*cusps*, i.e. for $\gamma > 0$, a function $f : \mathbb{R}^d \to \mathbb{R}$ has a γ-cusp at some point $\theta \in \mathbb{R}^d$, if there exists some finite $C > 0$ such that

$$|f^{(\lfloor\gamma\rfloor)}(\theta + \mathbf{h}) - f^{(\lfloor\gamma\rfloor)}(\theta)| \geq C||\mathbf{h}||^{\gamma - \lfloor\gamma\rfloor},$$

as $\mathbf{h}$ tends to zero. Another type of change-points are so-called *points of rapid change* considered by Menéndez et al. (2010), which are points with a rapid but smooth change occurring due to a local maximum of the first derivative with a certain magnitude.

In multivariate settings it is common to assume that the location of the change-points can be described by a curve, so that one rather speaks of a *change-curve*. We will sometimes abuse the denotations and simply refer to change-curves also as change-points for sake of brevity when speaking about change-point problems in nonparametric regression of arbitrary dimension.

Statistical challenges for change-point analysis

Initially, it is important to envision the numerous statistical issues arising in the context of NPP with change-points. Firstly, an obvious question is whether the unknown regression function is globally smooth or has some change-points, which naturally leads to estimation of the numbers of such change-points. This task is called *change-point-detection*. Secondly, given the existence of change-points, the explicit estimation of change-point-locations is of major interest. Thirdly, besides mere location of the change-point it is worthwhile to estimate the change-point-magnitude, such as the jump-height, to quantify the impact of the change-point on the considered function. Fourthly, having a reasonable believe that change-points are present and given reliable estimates for the three latter statistical tasks, the estimation of the (regression) function itself with incorporation of the change-points volunteers. The focus of this thesis lies strongly on the second issue and consequently the review of the literature will be mainly concentrated on this topic. For the other issues see the monograph by Qiu (2005) for further reading.

1.2.1 Estimation of change-point-locations in univariate frameworks

In the following we give a brief overview on the different methods, considered models and especially the coherences between the various frameworks in the literature about NPP with change-points. From now on we denote the regression function as m_θ to emphasize with the parameter θ that possibly change-points may be present.

Maximum-likelihood estimation

Ibragimov and Has' Minskii (1981) considered the regression function $m_\theta(x) = 1_{x\leq\theta}$ for some unknown $\theta \in (0,1)$ in the white noise model (*WNM*). This is a parametric regression model with a single jump-location θ. They showed that the maximum-likelihood estimate $\hat{\theta}_{MLE}$ of θ converges with the rate ϵ^2. In addition, it was shown that $\epsilon^{-2}(\hat{\theta}_{MLE} - \theta)$ converges in distribution with distributional limit

$$Z_0 = \arg\max_{u\in\mathbb{R}} B(u) - |u|/2, \tag{1.8}$$

where B is a two-sided Wiener process. For the related parametric regression problem with m_θ as before and different design assumptions, Korostelev and Tsybakov (1993) obtained the convergence rates of the least-squares estimate $\hat{\theta}_{LS}$ which coincides with the maximum-likelihood estimate for Gaussian errors. In case of equidistant design there is an asymptotically non-negligible bias of order $O(n^{-1})$, whereas for the random design the bias is asymptotically negligible. In the latter case the distributional limit of $n(\hat{\theta}_{LS} - \theta)$ is the discrete version of (1.8). Hinkley (1970) derived the asymptotic of the maximum-likelihood estimate by considering the mean-change problem in a sequence of Gaussian random variables which is basically the NPP for a fixed regular design and Gaussian errors. Korostelev (1987) verified for the jump-location estimation the minimax rate of convergence to be of order $O(\epsilon^2)$ in a WNM over a function class which is Lipschitz continuous away from the single discontinuity. By means of the usual correspondence this result can be transferred to a NPP which leads to a minimax rate of order $O(n^{-1})$. Thus, higher smoothness assumptions on the regression function away from the jump do not affect the minimax rate of convergence.

Smoothed difference approach

A typical assumption made in the univariate NPP with change-point is that the regression function (or its γ-th derivative) is of the form

$$m_\theta^{(\gamma)}(x) = m(x) + \sum_{i=1}^{k} a_i 1_{[\theta_i,\infty)}, \quad \gamma \in \mathbb{N} \cup \{0\}, \tag{1.9}$$

where $m : D \longrightarrow \mathbb{R}$ is a smooth function, $\theta = (\theta_1,\ldots,\theta_k) \in D^k$ are the change-point-locations and $a_i \neq 0$, $i = 1,\ldots,k$ the corresponding change-point-magnitudes. Note that for the case $k = 0$ the regression function (or its derivative) is smooth as well. Depending on this representation a popular method for estimating such change-point-locations in a univariate setting is to consider smoothed left- and right-sided limits of $m_\theta^{(\gamma)}$, say $\hat{m}_{n,-}^{(\gamma)}$ and $\hat{m}_{n,+}^{(\gamma)}$ and define an M-criterion function as the difference of both one-sided-smoothers. This method is called the *smoothed difference approach*. An estimate of the change-point-magnitude is naturally given by the value of the contrast function at its maximizer or minimizer. Note that the smoothed difference approach method volunteers for the detection of a change-point as well, by deciding that the change-point estimate is tagged as a change-point if the estimate of the change-point-magnitude exceeds a certain threshold (see Müller and Stadtmüller (1999) or Porter and Yu (2015) and their references for further reading). A related method based on weighted three smoothers, namely left- and right-sided smoothers and a central smoother, was considered by McDonald and Owen (1986) and Hall and Titterington (1992).
Most of the works using a smoothed difference approach differ by the assumptions on m, the smooth part of the regression function in (1.9), as well as the used one-sided smoothers, which will be pointed out in the following. Yin (1988) estimated change-points by means of a difference of one-sided averages in a WNM with unknown number of jumps. Qiu and Li (1991) considered in a NPP with fixed design differences of one sided-kernels as a contrast function to estimate jumps in the regression function, while Müller (1992) extended this method to estimate jump-point-locations and also kink-locations of higher order. In the same model with several jump-points, where the number of jumps is unknown, Wu and Chu (1993) used different types of kernels for the contrast function to estimate jump-location and -height as well as to give a procedure to estimate the number of jumps based on sequential hypothesis testing. Both latter papers obtained asymptotic normality of their jump-location estimate and their jump-magnitude estimate, while Müller (1992) also derived asymptotic normality of odd centered moments of his kink-location estimator. In Müller (1992) the one-sided kernels are supposed to vanish at zero, which in combination of an undersmoothing resulted in an unbiased asymptotic normality of the jump-location-estimates but with a sub-optimal rate of convergence and rather strong assumptions on the regression function. For weaker assumptions on the regression function, but with Gaussian errors and assuming that the change-point is one of the design points, Loader (1996) obtained the optimal rate by using differences of local polynomial smoothers as a criterion function and by letting the difference be strictly positive at zero. In order to derive the optimal rate of convergence Loader (1996) determined the asymptotic distribution of the estimate, which is a discrete version of (1.8). By assuming that the change-point is one of the design points the usually asymptotic non-negligible bias becomes negligible.
Based on a semi-parametric approach, which is asymptotically related to Müller (1992), Eubank and Speckman (1994) estimated a single kink-location and kink-magnitude of first order and obtained asymptotic normality of their estimates. Müller and Song (1997) and Gijbels et al. (1999) suggested two-stage estimation procedures which attain the optimal rate of convergence for the jump-location. In the first step, both defined a pilot estimate for an interval where the jump-location is located with high probability, while in the second step a refined estimate for the jump-location based on the first step is introduced by using a smoothed difference approach. Grégoire and Hamrouni (2002) considered the NPP with random design and used the difference of local linear smoother from left

and right as the criterion function. For a specific standardization of their M-estimate they obtained its asymptotic distribution, which is the maximizer of a compounded Poisson process, and also asymptotic normality of their jump-height estimate. Prieur (2007) obtained comparable results for a design with dependency structure, while Lin et al. (2008) also considered kink-location estimation of higher order and established asymptotic normality of exponents of their estimates.

Wavelet based methods

Another commonly used approach is based on *wavelet transformation* of the regression function as a criterion function. Wang (1995) suggested a wavelet transformation of the signal in a WNM with unknown number of γ-cusps of the underlying function to detect such cusps. The rate for estimation a γ-cusp with this technique is of order $O((\epsilon^2 \log(\epsilon)^{-\eta})^{1/(2\gamma+1)})$ for some $\eta > 0$. Raimondo (1998) derived the minimax rates of convergence for the estimation of the γ-cusp-location for function classes which have one single γ-cusp and are Hölder-smooth away from it. In addition, he provided a wavelet transformation based method to achieve his claimed minimax rate. He claimed that the minimax rate is of order $O(n^{-1/(2\gamma+1)})$, which is in fact true for $\gamma \in [0,1/2)$, but not for $\gamma \geq 1/2$ as shown by Goldenshluger et al. (2006).
Luan and Xie (2001), Park and Kim (2004) as well as Park and Kim (2006) considered basically the same estimates based on wavelet transformation as in Wang (1995) respectively Raimondo (1998) with different model assumptions for the NPP such as random design or some dependency structure in the errors and obtained the rate of convergence of order $O(n^{-1/(2\gamma+1)})$ for their cusp-location estimates. Furthermore, Huh and Carriere (2002) extended the approach of Loader (1996) to a kink of order γ estimation problem but without assuming Gaussian errors. They derived for their kink-location estimate the asymptotic distribution which is a generalized and discrete version of (1.8). However, all aforementioned procedures were out to attain the wrong minimax rates of Raimondo (1998) so that these methods are sub-optimal for $\gamma > 1/2$.

Zero-crossing time technique

A further popular method in the literature on image processing and change-point analysis is the so-called *zero-crossing-time-technique*. This name comes from using a Z-criterion function which is a smoothed second derivative of the object of interest. The motivation of this approach is that provided the function of interest has a jump-location, say at θ, a smoothed version of the function will have a large slope near θ, so that its first and second derivatives have a local maximum respectively a zero near θ.
Thus, for an appropriate kernel function $K : \mathbb{R} \to \mathbb{R}$, specified roughly below, the considered criterion function is

$$\psi_{h,m}(t) = h^{-(\gamma+1)} \int_0^1 K^{(\gamma+2)}\big(h^{-1}(x-t)\big)\, m(x)\, \mathrm{d}x, \tag{1.10}$$

where $h > 0$ is a bandwidth parameter and $\gamma \in \mathbb{N} \cup \{0\}$. Now, the kernel K is chosen such that $\psi_{h,m}$ has a well separated zero with a global maximum and a global minimum left respectively right from the zero and in addition, the zero of $\psi_{h,m}$ is close to the discontinuity of $m^{(\gamma)}$, see Figure 1.2 for illustration purposes. Chapter 3 provides more details about the particular choice of the kernel, as it is fundamental for the evolved theory in that chapter. An empirical Priestley-Chao-type of the criterion function is given by

$$\hat{\psi}_{n,h}(t) = n^{-1} h^{-(\gamma+1)} \sum_{i=1}^{n} Y_i\, K^{(\gamma+2)}(h^{-1}(x_i - t)). \tag{1.11}$$

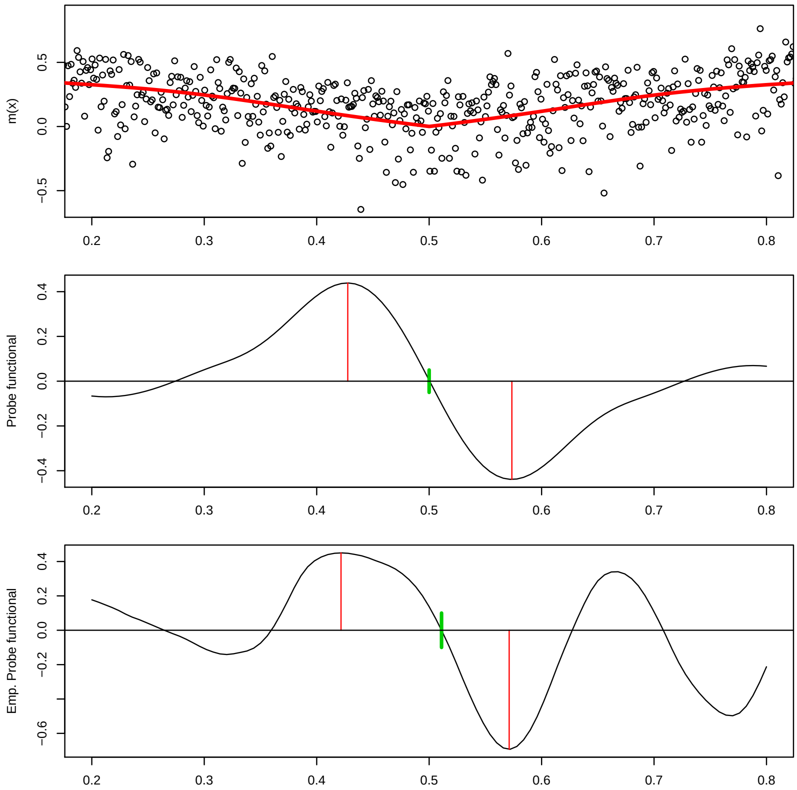

Figure 1.2.: Illustration of the zero-crossing-time-technique for a kink of first order. Top plot: Regression function with a single kink and noised observations. Middle plot: Corresponding criterion function in (1.10). Bottom plot: Empirical version in (1.11).

The advantage of this approach is that the global extreme values often allow a more accurate location of a change-point than other change-point-location methods.

Goldenshluger et al. (2006) estimated single jump-locations in an indirect WNM by using a zero-crossing-time-technique. Moreover, they derived the minimax lower bound for this estimation problem for Sobolev-smooth as well as analytical functions except for a single jump. This minimax lower bound can be related to the NPP with fixed design and in particular imply that the minimax rate of Raimondo (1998) is not correct for $\gamma \geq 1/2$, since the rate of convergence is driven by the smoothness of the regression function outside the cusp. Although their method is rate-optimal, it is not apparent how to apply the method for a direct nonparametric regression model with equidistant design. For this issue Cheng and Raimondo (2008) modified the method of Goldenshluger et al. (2006) by constructing kernels for their probe functional which is a smoothed version of the third derivative of the regression function for the purpose of kink-detection for kinks of first order. Goldenshluger et al. (2008a) and Goldenshluger et al. (2008b) embedded respectively extended the results of Goldenshluger et al. (2006) to a periodic setting and cover derivative estimation, convolution as well as delay and amplitude estimation, while Goldenshluger et al. (2008b) constructed adaptive estimators based on Lepski's adaption scheme.
The work of Wishart (2009) extended the method used in Cheng and Raimondo (2008) to the direct fractional-white-noise-model and the nonparametric regression model with equidistant design and dependency structure in the error, while Wishart (2011) established a minimax lower bound for this framework for slightly different function classes. For a random design in a similar framework, where the design follows some long-range dependency structure, Wishart and Kulik (2010) obtained rate of convergence which seem to be optimal.

Other methods

Finally, there are approaches worth noting which do not fit into the aforementioned categories, e.g. jump detection based on Fourier analysis was deployed by Lombard (1988). Dempfle and Stute (2002) considered the univariate NPP in a random design with very lax assumptions and defined a criterion function based on empirical quantiles to obtain an estimate which converges with the optimal rate.

1.2.2 Estimation of change-point-locations in multivariate frameworks

For the multivariate NPP with change-curves a prevalent model is the *boundary fragment model*, which is given by setting $D = [0,1]^d$ and

$$\begin{aligned} Y_{i_1,\dots,i_d} &= m_\phi(\mathbf{x}_{i_1,\dots,i_d}) + \varepsilon_{i_1,\dots,i_d}, \quad i_1,\dots,i_d \in \{1,\dots,n\}, \\ m_\phi(\mathbf{x}) &= m(\mathbf{x}) + \tau(\mathbf{x})\,1_{G(\phi)}, \\ G(\phi) &= \{(x_1,\dots,x_d) \in D \,|\, 0 \le x_d \le \phi(x_1,\dots,x_{d-1})\} \end{aligned} \tag{1.12}$$

where $m : D \to \mathbb{R}$ is a smooth curve, $\phi : [0,1]^{d-1} \to (0,1)$ the jump-location-curve and $\tau : D \to \mathbb{R}_+$ the corresponding jump-height-curve. This model is motivated by considering small parts of the original image (and further rescaling), where some smooth edge ϕ is located.
Korostelev and Tsybakov (1993) provided minimax results for jump-location-curves in Hölder-smoothness classes as well as a rate-optimal procedure for a random design based on piecewise-polynomial estimation on partitions of the design domain. The optimal rate of convergence depends only on the smoothness of the jump-location-curve and is not affected by additional smoothness of the smooth part m in (1.12), also referred to as a *nuisance parameter.* Moreover, for a regular fixed design with n^d design points, the minimax rate can not be faster than n^{-1} because of the dominating bias, so that the optimal rate is even not affected by additional smoothness assumptions on the jump-location-curve.

Rotated difference kernel estimation

Multivariate extensions of the idea of using differences of one-sided smoothers were introduced by Müller and Song (1994) and Qiu (1997) with the so-called *rotational difference kernel method.* This method allows to rotate the support of the corresponding one-sided kernels in order to adjust the criterion function such that its value is large near a change-curve. Garlipp and Müller (2007) pointed out that for the resulting estimates the order of scaling and rotating of the kernel support is important. We define the rotational difference kernel method for the bivariate version of the boundary fragment model in (1.12). Firstly, define the rotation matrix

$$\mathbf{D}_\psi = \begin{pmatrix} \cos(\psi) & \sin(-\psi) \\ \sin(\psi) & \cos(\psi) \end{pmatrix}, \qquad \psi \in \mathbb{R}. \tag{1.13}$$

Secondly, for a bandwidth $h > 0$ and $\mathbf{z} = (z_1, z_2)^T \in \mathbb{R}^2$ consider the rotated difference kernel

$$K(\mathbf{z};\psi,h) = K(h^{-1}\mathbf{D}_{-\psi}\mathbf{z})/h^2$$

where $K(\mathbf{z}) = K(z_1,z_2) = K_1(z_1)K_2(z_2)$ is a product kernel of univariate kernel functions K_1 and K_2, where K_2 is odd and justifies the term "difference". Finally, for $\mathbf{z} \in [0,1]^2$ and $\psi \in [-\pi/2, \pi/2]$ the contrast process with Priestley-Chao-type weights is defined as

$$\hat{\mathrm{M}}_n(\mathbf{z};\psi,h) = n^{-2} \sum_{i_1,i_2=1}^{n} Y_{i_1,i_2} K(\mathbf{z} - \mathbf{x}_{\mathbf{i_1},\mathbf{i_2}};\psi,h). \tag{1.14}$$

For an $x \in (0,1)$ denote by $\psi(x) = \arctan(\phi'(x))$ the slope of the tangent at $\phi(x)$. An estimator for the bivariate parameter $\big(\phi(x),\psi(x)\big)$ is then given by

$$\big(\hat{\phi}_n(x),\hat{\psi}_n(x)\big) \in \underset{y\in[h,1-h],\psi\in[-\pi/2,\pi/2]}{\arg\max} \hat{\mathrm{M}}_n((x,y)^T;\psi,h). \tag{1.15}$$

An illustration of the mechanism of the M-estimates in (1.15) is given in Figure 1.3. For some fixed x the largest contrast along the stripe $x \times (h, 1-h)$ is sought by rotating the kernel window appropriately. The line within the kernel window indicates along which direction the difference is considered. An estimator for the jump-height at x is given by

$$\hat{\tau}_n(x) = \hat{\mathrm{M}}_n((x,\hat{\phi}_n(x))^T;\hat{\psi}_n(x),h). \tag{1.16}$$

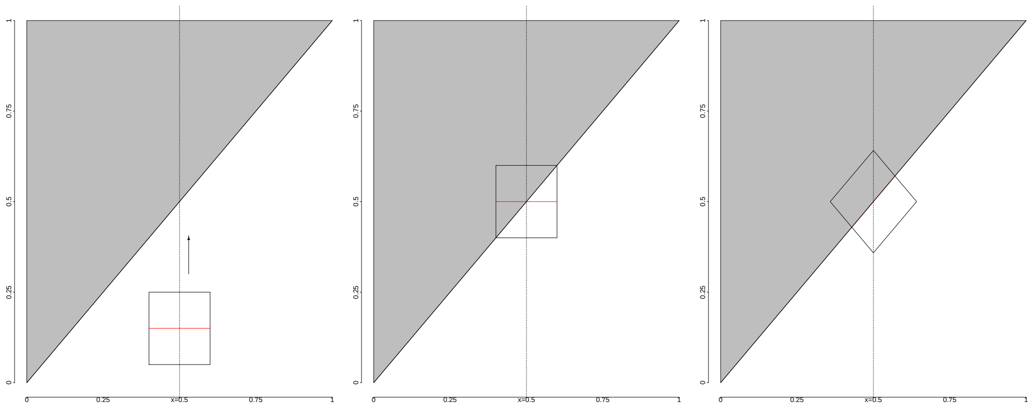

Figure 1.3.: Illustration of rotated difference kernel estimation for a linear jump-location-curve.

Other methods

A bivariate extension of the wavelet transformation technique of Wang (1995) for the NPP with equidistant design and Gaussian errors is given in Wang (1998). Another access to the multivariate NPP with change-curves is by treating the jump-location-curve as a pointset and consequently estimate it by a point set. This idea is motivated from developments in the image processing literature and the focus lies often on the detection of change-points. Qiu and Yandell (1997) suggested an algorithm to detect change-points by fitting local linear planes in the neighborhood of design points for the bivariate case, which was modified and extended to the three dimensional case in Mukherjee and Qiu (2011). Qiu (2002) proposed the simplified rotational difference kernel method for the detection of the jump-location-curve, while Sun and Qiu (2007) considered a criterion function based on estimation of first- and second-order-derivatives by local quadratic kernel smoothing. Moreover, Qiu (2002) showed that the rotational difference kernel can be related to the Sobel edge detector which is popular in the realm of image processing. For a review on the literature on image processing consider Qiu (2005) or Sonka et al. (2014).

1.3 Uniform M- and Z-estimation theory

As we have seen in the preceding section, many of the popular estimation procedures in the change-point analysis are simply M- or Z-estimates. In this section, M- and Z-estimation are introduced in two more general settings which firstly extend the classical theory for M- and Z-estimation and secondly will be highly relevant for the concepts in the further chapters. The setting is as general as possible and is therefore not restricted to the nonparametric regression framework.

1.3.1 Uniformity over the function class

Suppose the object of interest is a parameter θ_f of a function $f \in \mathcal{F}$ with

$$\mathcal{F} = \{f : I \to \mathbb{R} \,|\, f \text{ has some unique property in } \theta_f \in \Theta\},$$

where $\Theta \subset I$ is a compact subset and $I \subset \mathbb{R}^d$ is some set. In view of change-point problems it is evident that the parameter θ_f corresponds to a change-point. The considered statistical setting often allows to define for any $f \in \mathcal{F}$ a *(Z-)criterion function*

$$\begin{aligned} &\mathbb{R}^d \times \mathcal{F} \to \mathbb{R} \\ &(x, f) \mapsto \psi(x; f), \end{aligned}$$

such that the parameter of interest θ_f is a zero of $\psi(\cdot; f)$,. In order to estimate this parameter one might try as an estimate $\hat{\theta}_n$ a zero (provided it exists) of an *empirical (Z-)criterion function*

$$\begin{aligned} &\Omega \times \mathbb{R}^d \to \mathbb{R} \\ &(\omega, x) \mapsto \hat{\psi}_n(\omega, x), \end{aligned}$$

where we in the following suppress the dependency on ω and simply write $\hat{\psi}_n(x)$. We could also define an M-estimate in the same manner, but as we only need a Z-estimate in the further chapters we content ourselves with the explicit definition of Z-estimates in the above framework.

Asymptotic analysis

Assuming that the convergence of the empirical criteria function against the asymptotic criteria function holds uniformly in probability over the function class it is reasonable to believe that the corresponding Z-estimates converge to the actual parameter in probability and uniformly over the function class, provided the parameter of interest is somehow unique for the criterion function. Indeed, this heuristic is fruitful if the parameter θ_f is a well-separated zero, as the following proposition shows.

Proposition 1.1. *Let $\hat{\psi}_n : \mathbb{R}^d \to \mathbb{R}$ be random functions and let $\psi : \mathbb{R}^d \times \mathcal{F} \to \mathbb{R}$ be a deterministic function. Suppose that*

$$\sup_{\theta \in \Theta} |\hat{\psi}_n(\theta) - \psi(\theta; f)| = o_{P,\mathcal{F}}(1), \tag{1.17}$$

and that there exists a family $(\theta_f)_{f \in \mathcal{F}} \subset \Theta$ such that for any $\epsilon > 0$

$$\inf_{f \in \mathcal{F}} \inf_{\theta \in I: ||\theta - \theta_f||_\infty \geq \epsilon} |\psi(\theta; f)| > 0. \tag{1.18}$$

Then, for any estimator $\hat{\theta}_n$ with $\hat{\theta}_n \in \Theta$ and $\hat{\psi}_n(\hat{\theta}_n) = o_{P,\mathcal{F}}(1)$ it holds that

$$|\hat{\theta}_n - \theta_f| = o_{P,\mathcal{F}}(1).$$

Proof of Proposition 1.1. With (1.17) and the properties of $\hat{\theta}_n$ it follows that for any $\delta > 0$

$$\begin{aligned} P_f(|\psi(\hat{\theta}_n; f)| > \delta) &\leq P_f(|\hat{\psi}_n(\hat{\theta}_n)| > \delta/2) + P_f(|\hat{\psi}_n(\hat{\theta}_n) - \psi(\hat{\theta}_n; f)| > \delta/2) \\ &\leq o_{\mathcal{F}}(1) + P_f(\sup_{\theta\in\Theta} |\hat{\psi}_n(\theta) - \psi(\theta; f)| > \delta/2) = o_{\mathcal{F}}(1). \end{aligned}$$

Given $\epsilon > 0$ choose $\eta > 0$ as the left side of (1.18). Then,

$$P_f(|\hat{\theta}_n - \theta_f| > \epsilon) \leq P_f(|\psi(\hat{\theta}_n; f)| \geq \eta) = o_{\mathcal{F}}(1).$$

□

A remark on the conditions (1.17) and (1.18) is given in the succeeding subsection.

Having verified consistency of the estimates, a natural question arising is if $t_n(\hat{\theta}_n - \theta_f)$ converges in distribution to a non-degenerate limit for a suitable sequence of real-values $(t_n)_n \subset \mathbb{R}_+$. Fortunately, the nature of Z-estimates yields a convenient way for answering this question in case of a smooth criterion function. For sake of simplicity assume that $d = 1$, i.e. $\hat{\psi}_n$ is one-dimensional as well as that $\hat{\psi}_n$ is two-times differentiable and $\hat{\psi}_n^{(1)}$ is not zero in a certain neighborhood of θ_f, a Taylor expansion of $\hat{\psi}_n$ around θ_f yields

$$0 = \hat{\psi}_n(\hat{\theta}_n) = \hat{\psi}_n(\theta_f) + (\hat{\theta}_n - \theta_f)\hat{\psi}_n^{(1)}(\tilde{\theta}_{n,f}),$$

where $\tilde{\theta}_{n,f}$ is between θ_f and $\hat{\theta}_n$. Rearranging terms in the preceding display and multiplying by t_n leads to

$$t_n(\hat{\theta}_n - \theta_f) = -t_n\hat{\psi}_n(\theta_f)\left(\hat{\psi}_n^{(1)}(\tilde{\theta}_n)\right)^{-1}, \tag{1.19}$$

provided $\left(\hat{\psi}_n^{(1)}(\tilde{\theta}_n)\right)^{-1}$ exists. The term $t_n\hat{\psi}_n(\theta_f)$ is called the *(rescaled) score*. For a fixed function f it is possible to derive the asymptotic distribution as in the classical Z-estimation framework, see Section 5.3 in van der Vaart (2000). Though to obtain a stronger result, as for instance

$$\sup_{f\in\mathcal{F}} |P_f(t_n(\hat{\theta}_n - \theta_f) \leq \mathbf{x}) - F(\mathbf{x})| = o(1), \quad \forall \mathbf{x} \in \mathbb{R}^d$$

where F is some (known) cumulative distribution function, the results in part C of the appendix are helpful as well as Proposition 1.1. A specific application is given in Chapter 3.

1.3.2 Uniformity over the covariate space

In many statistical problems one has the following setting. $\Theta \subset \mathbb{R}^k$ is compact and $I \subset \mathbb{R}^d$ is some subset, the covariate space. Furthermore, one is interested in estimation of $f : I \to \Theta$ uniformly over the covariate space I. Fortunately, the statistical setting frequently allows to define a *(M-)criterion function*

$$\begin{aligned} &\Theta \times I \to \mathbb{R} \\ &(\theta, \mathbf{x}) \mapsto \mathbb{M}(\theta; \mathbf{x}) \end{aligned} \tag{1.20}$$

such that $f(\mathbf{x})$ is a maximizer or minimizer of $\mathbb{M}(\cdot;\mathbf{x})$ for any $\mathbf{x} \in I$. In order to estimate f one might try as an estimate $\hat{f}_n$ the maximizer or minimizer of an *empirical (M-)criterion function*

$$\begin{aligned}&\Omega \times \Theta \times I \to \mathbb{R}\\ &(\omega,\theta,\mathbf{x}) \mapsto \hat{\mathbb{M}}_n(\omega,\theta;\mathbf{x}),\end{aligned} \tag{1.21}$$

which converges pointwise in probability to $\mathbb{M}$. The estimate $\hat{f}_n(\mathbf{x})$ is called *M-estimator* in this framework. It is straightforward to define a Z-estimation in a similar setting, though this not of interest for this thesis and therefore omitted.

Asymptotic analysis

It is worth noting that the classical M-estimation approach as in van der Vaart (2000) is covered by the above setting as well. Indeed, set $f \equiv \theta_0$ and $\hat{\mathbb{M}}_n(\theta;\mathbf{x}) = \hat{\mathbb{M}}_n(\theta)$ respectively $\mathbb{M}(\theta;\mathbf{x}) = \mathbb{M}(\theta)$ for any $\mathbf{x} \in I$ in (1.20) and (1.21). Consequently, the pointwise consistency of an M-estimate $\hat{f}_n(\mathbf{x})$ for any $\mathbf{x} \in I$ can be shown by Theorem 5.7 in van der Vaart (2000) by fixing the $\mathbf{x}$-value.

For showing uniform consistency over I of an M-estimate we have the following extension of Theorem 5.7 in van der Vaart (2000), which also gives more flexibility to the sets of maximization/minimization Θ in (1.21).

Proposition 1.2. *Assume that $\tilde{\Theta}_{n,\mathbf{x}} \subset \mathbb{R}^k$ are compact sets and set $\Theta_n = \cup_{\mathbf{x}\in I}\{\mathbf{x}\} \times \tilde{\Theta}_{n,x}$. Let $\hat{\mathbb{M}}_n : \mathbb{R}^k \times I \to \mathbb{R}$ be random functions and let $\mathbb{M} : \mathbb{R}^k \times I \to \mathbb{R}$ be a deterministic function. Suppose that*

$$\sup_{(\mathbf{x},\theta)\in\Theta_n} |\hat{\mathbb{M}}_n(\theta;\mathbf{x}) - \mathbb{M}(\theta;\mathbf{x})| \xrightarrow{P} 0, \tag{1.22}$$

and that there exists a map $f : I \to \Theta$ such that for any $\epsilon > 0$

$$\inf_{\mathbf{x}\in I} \big(\mathbb{M}(f(\mathbf{x});\mathbf{x}) - \sup_{\theta\in\mathbb{R}^k : ||\theta - f(\mathbf{x})||_\infty \ge \epsilon} \mathbb{M}(\theta;\mathbf{x})\big) > 0. \tag{1.23}$$

Then for any estimator $\hat{f}_n(\mathbf{x}) \in \Theta_{n,\mathbf{x}}$, $\mathbf{x} \in I$, which satisfies $\inf_{\mathbf{x}\in I}\big(\hat{\mathbb{M}}_n(\hat{f}_n(\mathbf{x});\mathbf{x}) - \hat{\mathbb{M}}_n(f(\mathbf{x});\mathbf{x})\big) \ge -o_P(1)$ it holds that

$$||\hat{f}_n - f||_\infty \xrightarrow{P} 0,$$

where we assume that $||\hat{f}_n - f||_\infty$ is measurable.

Proof of Proposition 1.2. From (1.22) we have that $\sup_{\mathbf{x}\in I} |\hat{\mathbb{M}}_n(f(\mathbf{x});\mathbf{x}) - \mathbb{M}(f(\mathbf{x});\mathbf{x})| = o_P(1)$. By the property of the estimator $\hat{f}_n(\mathbf{x})$ it follows that $\sup_{\mathbf{x}\in I}\big(\hat{\mathbb{M}}_n(f(\mathbf{x});\mathbf{x}) - \hat{\mathbb{M}}_n(\hat{f}_n(\mathbf{x});\mathbf{x})\big) \le o_P(1)$. With this, $\sup_{\mathbf{x}\in I}\big(\mathbb{M}(f(\mathbf{x});\mathbf{x}) - \hat{\mathbb{M}}_n(\hat{f}_n(\mathbf{x});\mathbf{x})\big) \le o_P(1)$. Next, using (1.22) again yields

$$\begin{aligned}\sup_{\mathbf{x}\in I}\big(\mathbb{M}(f(\mathbf{x});\mathbf{x})\big) - \mathbb{M}(\hat{f}_n(\mathbf{x});\mathbf{x}) &\le \sup_{\mathbf{x}\in I}\big(\mathbb{M}(f(\mathbf{x});\mathbf{x}) - \hat{\mathbb{M}}_n(\hat{f}_n(\mathbf{x});\mathbf{x})\big) + \sup_{\mathbf{x}\in I}\big(\hat{\mathbb{M}}_n(\hat{f}_n(\mathbf{x});\mathbf{x}) - \mathbb{M}(\hat{f}_n(\mathbf{x});\mathbf{x})\big)\\ &\le o_P(1) + \sup_{(\mathbf{x},\theta)\in\Theta_n} |\hat{\mathbb{M}}_n(\theta;\mathbf{x}) - \mathbb{M}(\theta;\mathbf{x})| = o_P(1).\end{aligned}$$

Given $\epsilon > 0$ choose $\eta > 0$ as the left side of (1.23). Then

$$\begin{aligned}P\big(||\hat{f}_n(\mathbf{x}) - f(\mathbf{x})||_\infty > \epsilon\big) &= P\big(\exists\, \mathbf{x} \in I : |(\hat{f}_n(\mathbf{x}) - f(\mathbf{x}))_i| \ge \epsilon\big)\\ &\le P\big(\exists\, \mathbf{x} \in I : \mathbb{M}(f(\mathbf{x});\mathbf{x}) - \mathbb{M}(\hat{f}_n(\mathbf{x});\mathbf{x}) \ge \eta\big)\\ &= P\big(\sup_{\mathbf{x}\in I}\big(\mathbb{M}(f(\mathbf{x});\mathbf{x}) - \mathbb{M}(\hat{f}_n(\mathbf{x});\mathbf{x})\big) \ge \eta\big) = o(1).\end{aligned}$$

$\square$

Concentration inequalities or large deviation inequalities are powerful tools to verify the conditions in (1.17) or (1.22), while the conditions (1.18) and (1.23) require a rigorous analysis of the corresponding limit criterion function.

As in the scenario before, one may ask if $t_n(\hat{f}_n - f)$ converges in distribution to a non-degenerate limit for a suitable sequence of real-values $(t_n)_n \subset \mathbb{R}_+$. Pointwise, that is for any fixed $\mathbf{x} \in I$, the classical approach can be employed to obtain that

$$t_n(\hat{f}_n(\mathbf{x}) - f(\mathbf{x})) = -t_n \nabla \hat{\mathbb{M}}_n(f(\mathbf{x});\mathbf{x})\big(\nabla\nabla^T \hat{\mathbb{M}}_n(\tilde{\theta}_n(\mathbf{x});\mathbf{x})\big)^{-1}, \tag{1.24}$$

for some $\tilde{\theta}_n(\mathbf{x})$ between $f(\mathbf{x})$ and $\hat{f}_n(\mathbf{x})$ and provided $\big(\nabla\nabla^T \hat{\mathbb{M}}_n(\tilde{\theta}_n(\mathbf{x});\mathbf{x})\big)^{-1}$ exists. Thus, the asymptotics of $t_n(\hat{f}_n(\mathbf{x}) - f(\mathbf{x})$ for arbitrary points $\mathbf{x}$ can be investigated similar to the classical case. Moreover, equation (1.24) is a comfortable expression to analyze the asymptotic behavior of $||\, t_n(\hat{f}_n - f)\,||_\infty$, provided the score and the Hessian matrix of this statistic possess an appropriate asymptotic behavior or this statistic can be suitable approximated, see for instance Example 1 below and the discussion before.

1.3.3 Distinct rates of convergence

In order to guarantee an appropriate limit distribution for the terms in (1.19) resp. (1.24) the components of the estimates should have all the same rate of convergence. For this to hold, an appropriate criterion function must be used, as for instance a dilated criterion function $\hat{\mathrm{W}}_n(\mathbf{w}) = \hat{\psi}_n(\theta_f + \mathbf{s}_n \circ \mathbf{w})$, where $\circ : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ is the Hadamard product and $\mathbf{s}_n \subset \mathbb{R}^d_+$ is some sequence normalizing the speed of convergence such that the components of the maximizer $\hat{\mathbf{w}}_n$ of $\hat{\mathrm{W}}_n(\mathbf{w})$ have all the same rate of convergence. See Remark 1 in Chapter 2 for an application in the setting of Section 1.3.2.

1.4 Confidence sets

One of the challenging statistical tasks is to construct sets, so-called *confidence sets*, based on the sample such that an object of interest is contained in this set with a pre-defined high probability, while these sets should be as small as possible. Apparently, the construction of confidence sets requires a reliable estimate of the object of interest and a sufficient knowledge about the uncertainty of the estimate in order to satisfy both requirements on the confidence sets.
Knowledge about the uncertainty of the estimate can be gained by investigating its finite or asymptotic distribution, while the reliability of an estimate from a statistician view is captured by risk-functions, introduced in Section 1.1. These risk-functions are besides the parameter space $\mathcal{F}$ mainly driven by the considered metric d. In this thesis, only the risks induced by the sup-norm are investigated more further. For construction of confidence sets based on the L_2-norm see Robins and van der Vaart (2006) and the references therein.
Let $I \subset \mathbb{R}^d$. Suppose our object of interest is a function $f : I \to \mathbb{R}^k$ which is element of a function class $\mathcal{F}$. For any $\mathbf{x} \in I$ and $\alpha \in (0,1)$ let $C_n(\mathbf{x},\alpha) \subset \mathbb{R}^k$ be a random subset. If for any $\alpha \in (0,1)$ holds

$$P_f(f(\mathbf{x}) \in C_n(\mathbf{x},\alpha)\ \forall \mathbf{x} \in I) \geq 1 - \alpha, \tag{1.25}$$

then $C_n(\alpha) = (C_n(\mathbf{x},\alpha))_{\mathbf{x}\in I}$ is called a *level* $(1-\alpha)$ *confidence band* for f. If the constructed confidence band is such that inequality (1.25) holds for the limit case n to infinity, the confidence band is called *asymptotic level* $(1-\alpha)$ *confidence band* for f. In particular, if f is a constant function and $k = 1$, then one uses rather the label *confidence interval* for C_n. The *width* of the confidence set $C_n(\alpha)$ is given by $w(\alpha) = \sup_{x\in I} \lambda_k(C_n(\mathbf{x};\alpha))$.

Clearly, the condition in (1.25) is not stringent as one can choose a large subset of $\mathbb{R}^k$ and the inequality is certainly fulfilled. Hence, a sensible condition to hold is that the width converges to zero for n tending to infinity for any fixed $\alpha \in (0,1)$. In other words, the more observations are available the more the confidence set clings to f.
For illustration purposes consider Figure 1.4, where some valid confidence bands are plotted for the function $f(x) = \sin(\pi x)$. Apparently, the confidence band in the left picture does not give much information about the behavior of the function, while the confidence band on the right provides more information about some properties of the function such as monotonicity along some subintervals or possible locations of extreme values. Therefore, the requirements on the confidence band are on the one hand that (1.25) is a sharp lower bound, on the other hand the width should be as small as possible. Obviously there is a trade-off between both requirements which is comparable to the

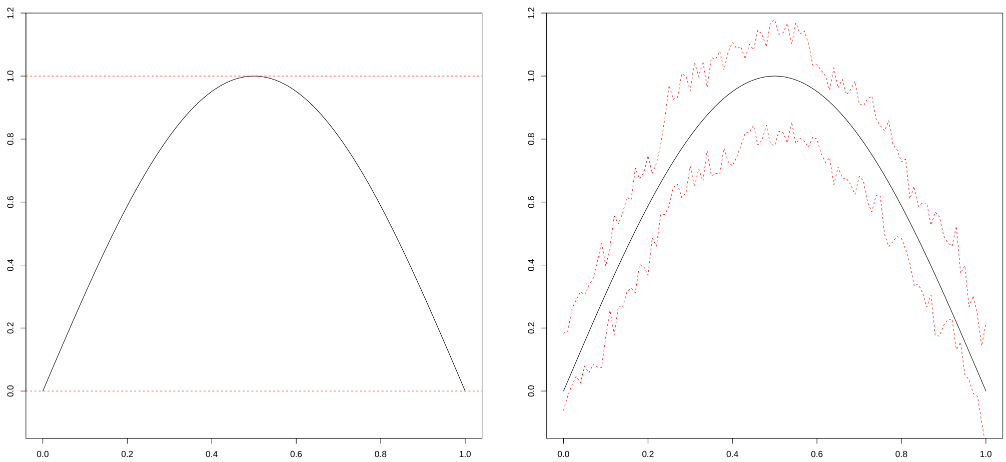

Figure 1.4.: Left: Bad choice for a confidence band. Right: Sensible choice for a confidence band.

bias-variance trade-off emerging in nonparametric estimation problems, i.e. making the confidence band slender reduces its variation but makes it susceptible for deviation from the function of interest and thus violating condition (1.25) and vice versa.
An appealing technique to construct asymptotic confidence bands is to define a uniform consistent estimate $\hat{f}_n$ of f and use it as the center of the confidence bands. To adjust the width of the confidence band, one might deduce a non-degenerated distributional limit of $||t_n\Sigma_n^{-1/2}(\hat{f}_n - f)||_\infty$ for some appropriate sequence of real-values $(t_n)_n \subset \mathbb{R}_+$ and $\Sigma_n(\mathbf{x})$ the covariance matrix of $t_n\hat{f}_n(\mathbf{x})$. Assume that a non-degenerated distributional limit exists and denote its distribution by Q. Furthermore, let $\hat{\Sigma}_n$ be a uniform consistent estimate for Σ_n, then for any $\alpha \in (0,1)$ an asymptotic level $(1-\alpha)$ confidence band for f is given by

$$C_n(\alpha) = \left\{[c_f^-(\mathbf{x};\alpha), c_f^+(\mathbf{x};\alpha)] \mid \mathbf{x} \in I\right\}, \quad c_f^{\pm}(\mathbf{x};\alpha) = \hat{f}(\mathbf{x}) \pm \frac{\hat{\Sigma}_n^{1/2}(\mathbf{x})\, q_{1-\alpha}(Q)}{t_n}. \tag{1.26}$$

Indeed,

$$\begin{aligned} &P_f\big(f(\mathbf{x}) \in [c_f^-(\mathbf{x};\alpha), c_f^+(\mathbf{x};\alpha)] \quad \forall \mathbf{x} \in I\big) \\ &\qquad = P_f\big(||t_n\hat{\Sigma}_n^{-1/2}(\hat{f}_n - f)||_\infty \le q_{1-\alpha}(Q)\big) \\ &\qquad \approx P_f\big(||t_n\Sigma_n^{-1/2}(\hat{f}_n - f)||_\infty \le q_{1-\alpha}(Q)\big) + o(1) \to 1-\alpha. \end{aligned}$$

Closer consideration of the preceding display suggests to use the quantiles of $||t_n\Sigma_n^{-1/2}(\hat{f}_n - f)||_\infty$, say $c(\alpha)$. However, this approach is in general not satisfactory for the nonparametric case as the quantiles $c(\alpha)$ are generally not feasible. This is because of the usually unknown finite distribution of $||t_n\Sigma_n^{-1/2}(\hat{f}_n - f)||_\infty$. Instead, a common technique is to find a centered Gaussian process $G_{n,f}$ with estimable covariance structure such that $||t_n\Sigma_n^{-1/2}(\hat{f}_n - f)||_\infty$ and $||G_{n,f}||_\infty$ are close in some appropriate sense. Consequently, the quantiles of $||G_{n,f}||_\infty$ serve as approximations for the desired quantiles $c(\alpha)$. This method is called *Gaussian approximation* and was first introduced by Smirnov (1950) for histogram-based estimates and by Bickel and Rosenblatt (1973) for kernel-based estimators.
There are mainly two approaches to determine or rather approximate the quantiles of $||G_{n,f}||_\infty$, which can be roughly categorized into an analytical method and a simulation method. The analytical method is based on the so-called *Smirnov-Bickel-Rosenblatt-condition*, which is showing that $||G_{n,f}||_\infty$ converges in distribution to a Gumbel distribution if suitably translated and scaled. Recently, Chernozhukov et al. (2014) provided a generalized Smirnov-Bickel-Rosenblatt-condition making this access somewhat easier. The simulation method uses stochastic simulation methods such as a multiplier bootstrap to compute $q_{1-\alpha}(||G_{n,f}||_\infty)$.
In summary, the resulting confidence bands based on Gaussian approximation are of the form

$$C_n(\alpha) = \left\{[c^-(\mathbf{x};\alpha), c^+(\mathbf{x};\alpha)] \mid \mathbf{x} \in I\right\}, \quad c^\pm(\mathbf{x};\alpha) = \hat{f}(\mathbf{x}) \pm \frac{\hat{\Sigma}_n(\mathbf{x})\, c_n(1-\alpha)}{t_n}, \tag{1.27}$$

where $c_n(1-\alpha)$ is possibly an approximation for the quantile $q_{1-\alpha}(||G_{n,f}||_\infty)$. The following example illustrates the construction of confidence sets based on the Smirnov-Bickel-Rosenblatt-condition for the Nadaraya-Watson-estimate.

Example 1. Let us assume for model (1.2) that $m \in \mathcal{H}^2([0,1], L)$ for some finite constant $L > 0$. Furthermore, let the design be equidistant and let the errors ε_i be such that $\mathbb{E}|\varepsilon_i|^5 < \infty$. We investigate the Nadaraya-Watson estimate (see for instance Tsybakov (2009)) as a Z-estimate, by defining the Z-criterion function

$$\hat{\psi}_n(\theta; x) = \frac{1}{nh}\sum_{i=1}^n (Y_i - \theta) K\big((x_i - x)/h\big),$$

where $K : \mathbb{R} \to \mathbb{R}_+$ is a continuous differentiable function with $\mathrm{supp}(K) = [-1,1]$ and $\int K = 1$ as well as $\int xK(x)\,\mathrm{d}x = 0$. Moreover, the bandwidth $h = h_n$ is such that $h \in (C_1 n^{-1/3}, C_2 n^{-1/5})$ for some finite constants $C_1, C_2 > 0$. We denote the zero of $\hat{\psi}_n(\cdot; x)$ by $\hat{m}_n(x)$ for any $x \in [0,1]$. In addition, we let $I \subset (0,1)$ be a compact subset and assume that n is large enough such that $I \subset [h, 1-h]$. Then, in the spirit of (1.19) for any $x \in (0,1)$

$$\hat{m}_n(x) - m(x) = -\hat{\psi}_n(m(x); x)\left(\hat{\psi}_n^{(1)}(\tilde{m}(x); x)\right)^{-1},$$

for some $\tilde{m}(x)$ between $m(x)$ and $\hat{m}_n(x)$. Obtain by Lemma B.4 in the appendix that

$$\begin{aligned} H_n(x) := \hat{\psi}_n^{(1)}(\tilde{m}(x); x) &= -\frac{1}{nh}\sum_{i=1}^n K\big((x_i - x)/h\big) \\ &= -\int K(x)\,\mathrm{d}x + O_{x\in[0,1]}((nh)^{-1}) = -1 + O_{x\in[0,1]}((nh)^{-1}). \end{aligned} \tag{1.28}$$

Additionally, define

$$\begin{aligned} \sqrt{nh}\,\hat{\psi}_n(m(x); x) &= \frac{1}{\sqrt{nh}}\sum_{i=1}^n \varepsilon_i K\big((x_i - x)/h\big) + \frac{1}{\sqrt{nh}}\sum_{i=1}^n (m(x_i) - m(x)) K\big((x_i - x)/h\big) \\ &=: Z_n(x) + R_n(x). \end{aligned}$$

Lemma B.5 in the appendix implies that

$$\mathbb{E}\sup_{x\in I}|Z_n(x)| \le C\sqrt{\log(n)}, \tag{1.29}$$

for some finite constant $C > 0$. By Taylor expansion of m

$$m(x+yh)-m(x) = yhm^{(1)}(\tilde{x}),$$

where $\tilde{x} = \rho x + (1-\rho)(x+yh)$ for some $\rho \in [0,1]$ is a value between x and $x+yh$. Hence, by Lemma B.4 in the appendix and Lipschitz continuity of m

$$\begin{aligned}\frac{|R_n(x)|}{\sqrt{nh}} &= \Big|\int \big(m(x+yh)-m(x)\big)K(y)\,\mathrm{d}y\Big| + O_{x\in[0,1]}\big((nh)^{-1}\big)\\ &= h\Big|\int ym^{(1)}(\tilde{x})K(y)\,\mathrm{d}y\Big| + O_{x\in[0,1]}\big((nh)^{-1}\big)\\ &= h\Big|\int y\big(m^{(1)}(\tilde{x})-m^{(1)}(x)\big)K(y)\,\mathrm{d}y\Big| + O_{x\in[0,1]}\big((nh)^{-1}\big)\\ &\le Lh^2\int |y|^2|K(y)|\,\mathrm{d}y + O_{x\in[0,1]}\big((nh)^{-1}\big),\end{aligned} \tag{1.30}$$

where we used the vanishing first moment of K for the third equation and Lipschitz continuity of $m^{(1)}$ in the last line. In summary,

$$\sqrt{nh}\big(\hat{m}_n(x)-m(x)\big) = -\big(Z_n(x)+R_n(x)\big)/H_n(x), \tag{1.31}$$

where R_n is a bias term and Z_n is a stochastic term or sometimes also called variance term. For the construction of confidence bands we need to approximate $\sup_{x\in I}|Z_n(x)|$, as the other terms R_n and H_n are deterministic, while R_n can be made asymptotically negligible by choosing the bandwidth h appropriately.

It can be shown that under the assumptions on ε_i and K that for any fixed $x \in (0,1)$ it holds $Z_n(x) \rightsquigarrow N(0, \int K^2(t)\,\mathrm{d}t)$ (see for instance Lemma 2.10). Hence, it seems reasonable to approximate $\sup_{x\in I}|Z_n(x)|$ by the supremum of some Gaussian process. For this purpose, obtain by Lemma A.2 in the appendix that there exists a Brownian motion W on $\mathbb{R}$ (by probably enriching the probability space) such that

$$\sup_{x\in I}|Z_n(x)-Z_G(x)| = O_P\Big(\sqrt{\frac{\log(n)}{nh}}\Big), \tag{1.32}$$

where $Z_G(x) = h^{-1/2}\int K(h^{-1}(x-y))\,dW(y)$. This is a centered Gaussian process with standard deviation $\sigma(K) = \big(\int K^2(x)\,\mathrm{d}x\big)^{1/2}$, which can be seen by Itô-Isometry. In Härdle (1989) one can find the following result, based on Bickel and Rosenblatt (1973): There exists a finite constant $C > 0$ such that for $d_n = C\sqrt{\log(n)}$, and $\delta \in (1/5, 1/3)$, it holds for any $y \in \mathbb{R}$

$$P\Big(\sqrt{2\delta\log(n)}\big(\sup_{x\in I}{}^{|Z_G(x)|}/_{\sigma(K)} - d_n\Big) \le y\big) \to \exp(-2\exp(-y)). \tag{1.33}$$

Note that this convergence corresponds to the classical Smirnov-Bickel-Rosenblatt-condition, as the right-hand side of the latter display is the cumulative distribution function of a standard Gumbel distribution. With these results it is straightforward to define confidence bands for $m(x)$ by

$$c^{\pm}(x;\alpha) := \hat{m}_n(x) \pm \frac{\sigma(K)\,c_n(\alpha)}{\sqrt{nh}}, \tag{1.34}$$

where

$$c_n(\alpha) = (1 - v_n)\Big(d_n + \frac{\log(2) - \log|\log(1-\alpha)|}{\sqrt{2\delta\log(n)}} + 2b_n\Big)$$

and $b_n \cong \sqrt{\log(n)}$ such that

$$P\big(\sup_{x\in I}|R_n(x)| \geq b_n\,\sigma(K)\big) = o(1),$$

which can be achieved by an undersmoothing, i.e. $h^2 < \sqrt{\frac{\log(n)}{nh}}$ for n sufficiently large (compare the asymptotics in (1.29) and (1.30)). In addition, due to (1.32),

$$P\big(\sup_{x\in I}|Z_n(x) - Z_G(x)| \geq b_n\,\sigma(K)\big) = o(1).$$

Moreover, $v_n \cong (nh)^{-1}$ is such that

$$P\big(\sup_{x\in I}|H_n^{-1}(x)| \geq (1-v_n)^{-1}\,\sigma(K)\big) = o(1),$$

compare to (1.28). Thus, by (1.31)

$$\begin{aligned}
&P(m(x) \notin [c^-(x;\alpha), c^+(x;\alpha)]\ \forall x \in I)\\
&\quad= P\big(\sup_{x\in I}\sqrt{nh}\,|m(x)-\hat{m}_n(x)|/\sigma(K) \geq c_n(\alpha)\big)\\
&\quad\leq P\big(\sup_{x\in I}|Z_n(x) - Z_G(x)| \geq b_n\,\sigma(K)\big) + P\big(\sup_{x\in I}|Z_G(x)|/\sigma(K) \geq c_n(\alpha)(1-v_n)^{-1} - 2b_n\big)\\
&\qquad+ P\big(\sup_{x\in I}|R_n(x)| \geq b_n\,\sigma(K)\big) + P\big(\sup_{x\in I}|H_n^{-1}(x)| \geq (1-v_n)^{-1}\,\sigma(K)\big)\\
&\quad= \alpha + o(1),
\end{aligned}$$

where we used (1.33) in the last line. Thus, (1.34) yields an asymptotic level $(1-\alpha)$ confidence band and its width is of order $\sqrt{\log(n)/nh}$. In view of minimax optimality results over Hölder function classes as in Tsybakov (2009), the width is of a slightly larger order as the optimal rate of convergence due to undersmoothing. ■

A certainly relevant aspect of nonparametric estimation is that nonparametric estimators are biased, that is the expected value of the estimate based on finite data deviates from the object of interest. As a result, it is crucial to take the bias into account in order to ensure an appropriate center of the confidence bands based on the approaches above.

There are two common ways to cope with the bias for the confidence bands construction. One way is by estimating the bias explicitly and to correct the center of the confidence sets by this bias-estimate, in which case the confidence bands are called *bias-corrected.* The other way is based on the bias-variance trade-off emerging in nonparametric estimation by choosing the tuning parameters of the nonparametric estimate such that the bias is asymptotically negligible compared to the variance. In other words, an undersmoothing is employed to deal with the bias (see Example 1). Concerning the coverage property of confidence bands, Hall (1992) and Neumann (1995) showed that undersmoothing is preferable to bias-correction in many situations. In view of these results, it has become common practice to use an undersmoothing for confidence bands construction, even though this comes at the expanse of slightly slower rates of convergence resp. marginally wider confidence bands.

Confidence bands for smooth functions f are by now well-developed in the density estimation, see e.g. Bickel and Rosenblatt (1973), Neumann (1998), Giné and Nickl (2010), Hoffmann and Nickl (2011) and Chernozhukov et al. (2014). The confidence band construction for smooth regression function in the NPP had been studied rigorously by Härdle and Bowman (1988), Härdle (1989), Eubank and Speckman (1993) or Neumann and Polzehl (1998). For inverse problems for density and regression function estimation, there are works by Bissantz and Holzmann (2008), Birke et al. (2010) and Lounici and Nickl (2011). Moreover, Proksch (2016) considered a multivariate nonparametric regression framework, while Proksch et al. (2015) investigated a multivariate deconvolution setting. More recently, Mammen and Polonik (2013) and Qiao and Polonik (2016) focused on more geometrical features and constructed confidence sets for density level sets and the density ridge, respectively. For a nice textbook introduction and several extensions on the construction of confidence bands see Giné and Nickl (2015).

Adaptive confidence sets

As pointed out in Section 1.1 it is common practice in nonparametric statistics to assume that the object of interest f is element of a function class $\mathcal{F} = \mathcal{F}_s$ endowed with some parameter $s > 0$ controlling the smoothness of the functions in some sense, such that the (minimax-optimal) rate of convergence r_n frequently depends on the smoothness parameter s, say $r_n(s)$. As a consequence, suitable confidence sets as in (1.26) or (1.27) depend on the smoothness parameter s as well in order to guarantee a width of nearly the same order as the optimal rate of convergence.
However, the a-priori knowledge about the smoothness of an unknown function is a strong limiting assumption so that it is preferable to define an estimate which adapts to the smoothness of the function automatically. For this purpose, some data-driven methods have been developed which achieve the minimax-optimal rate of convergence (up to a logarithmic factor in some cases) by only assuming to have some candidate set $\mathcal{S} \subset \mathbb{R}_+$ which contains the actual smoothness level s of f, i.e. $f \in \mathcal{F} = \bigcup_{s\in\mathcal{S}} \mathcal{F}_s$. To manifest such a property into a definition, an estimate $\hat{f}_n$ is called *adaptive* over $(\mathcal{F}, d)$ if

$$\sup_{s\in\mathcal{S}} R_w\left(\hat{f}_n; \mathcal{F}_s, d, \tilde{r}_n(s)\right) \leq C,$$

where $\tilde{r}_n(s)$ is the minimax-optimal rate of convergence $r_n(s)$ for $(\mathcal{F}_s, d)$ possibly inflated by a logarithmic term in n and $C > 0$ is some finite constant.
One popular method for adaptive estimation is due to Lepskii (1992) which is often referred to as *Lepski's scheme* or *Lepski's method.* Roughly speaking, this method chooses the tuning parameter of the nonparametric estimation procedure in such a way that the probability of an oversmoothing or an undersmoothing tends to zero. There are various frameworks in which Lepski's scheme had been successfully applied, for instance: Tsybakov (1998), Butucea (2001), Giné and Nickl (2009), Giné and Nickl (2010) or Gach et al. (2013).
The next obvious question is how the adaptivity of the estimates can be used to construct confidence bands which adapt to the smoothness of each individual $f \in \mathcal{F} = \bigcup_{s\in\mathcal{S}} \mathcal{F}_s$. The most common view on adaptivity of a confidence band is due to Cai and Low (2004), which declares a confidence band $C_n(\alpha)$ *adaptive for the class* $\mathcal{F} = \bigcup_{s\in\mathcal{S}} \mathcal{F}_s$, if for any $s \in \mathcal{S}$ and any $\epsilon > 0$ there exists a finite constant $C = C(\alpha) > 0$ depending only on $\alpha \in (0,1)$ such that

$$\sup_{f\in\mathcal{F}_s} P_f\left(w(\alpha) \geq C(\alpha)\tilde{r}_n(s)\right) < \epsilon \quad \forall \alpha \in (0,1), \tag{1.35}$$

where $\tilde{r}_n(s)$ equals the minimax-optimal rate of convergence $r_n(s)$ possibly up to a logarithmic factor in n. An alternative approach to define adaptivity of confidence bands is suggested by Genovese and Wasserman (2008). In addition to the adaptivity in (1.35), the confidence bands should maintain appropriate coverage properties over the considered function class $\mathcal{F}$. As a result, we call a confidence band *honest with level α for the class $\mathcal{F}$*, if

$$\inf_{f \in \mathcal{F}} P_f\big(f(\mathbf{x}) \in C_n(\mathbf{x},\alpha)\ \forall \mathbf{x} \in I\big) \geq 1 - \alpha. \tag{1.36}$$

If the latter inequality holds for taking the limes inferior over n, the confidence bands are called *asymptotically honest with level α for the class $\mathcal{F}$*. These definitions are due to Li (1989).
Unfortunately, there are several negative results for the construction of honest and at the same time adaptive confidence bands for various classes and nonparametric frameworks (Low, 1997; Cai and Low, 2004; Genovese and Wasserman, 2008). These negative results can be summarized to: while adaptive estimation can be employed in specific frameworks, it is in general not possible in these cases to construct honest and adaptive confidence bands. Nevertheless, there have been works which showed that under certain conditions it is indeed perfectly possible to construct adaptive and honest confidence sets based on adaptive estimation given by Lepski's method (Picard and Tribouley, 2000; Giné and Nickl, 2010; Bull, 2012; Chernozhukov et al., 2014), which however is not the only option (Dümbgen, 2003; Cai et al., 2013). All these methods have in common that further conditions on $\mathcal{F}$, such as shape constraints or the so-called self-similarity (see Bull (2012) or Giné and Nickl (2015)) are required in order to proceed.

1.5 Confidence sets in nonparametric regression with change-points

Evidently, objects of interest in the NPP with change-points are the characteristics of change-points, for instance in the univariate case the change-point-location or in the multivariate case the jump-location-curve. In the following we give a review on the literature concerning the construction of confidence sets for jump-discontinuities of the regression function as well as for kink-locations of higher order.

Confidence sets for jump-locations

One of the first nonparametric methods for constructing confidence sets for change-points in a sequence of independent identical distributed observations is by Dümbgen (1991) based on a bootstrap procedure. The works of Müller (1992) and Loader (1996) provided techniques to obtain asymptotic confidence sets for a jump-point-location in a univariate NPP, although only Loader (1996) constructed explicitly confidence sets in his work. However, both approaches have drawbacks considering the construction of confidence sets, as on the one hand the estimate of Müller (1992) is asymptotically normal, on the other hand his estimate does not achieve the optimal rate, while the estimate in Loader (1996) is rate optimal, though it has a non-standard asymptotic distribution.
A rate optimal bootstrap algorithm to construct confidence sets for jump-point-locations as well as the regression function itself, which is based on the detection method of Gijbels et al. (1999), was suggested by Gijbels et al. (2004). For a stochastic design and a two-phase linear regression function Seijo and Sen (2011) provided a bootstrap-based method for the construction of confidence intervals of a jump-location with optimal rate of convergence.
However, these papers considered all one-dimensional models and the methods are not applicable in cases where the design is multivariate. As pointed out in Proksch et al. (2015), there are considerable distinctions between the construction of confidence sets in a univariate setting and a multivariate setting. In addition, there seems to be no methods available to construct confidence sets for the jump-location-curve for multivariate cases.
To close this gap, Chapter 2, which is based on Bengs et al. (2018), provides a method to construct confidence sets for the jump-location-curve based on the rotational difference kernel method by Qiu (1997) or Müller and Song (1994).

Confidence sets for kink-locations

Although much effort was invested for estimation of kink-locations (see Section 1.2.1), the literature is sparse concerning the construction of asymptotic confidence intervals for kink-locations, especially for estimates which attain the optimal rate of convergence. The approaches of Müller (1992), Eubank and Speckman (1994), Huh and Carriere (2002) offer possibilities to obtain asymptotic confidence intervals for kinks based on the asymptotic normality of their estimates. Even though the results of Müller (1992) are applicable for higher order kinks, the resulting confidence intervals tend to be wider as the asymptotic normality holds only for odd exponents of his kink estimate. In addition, the assumptions on the regression function are rather strict.

As already mentioned in Section 1.2.1, the rates of convergence in Huh and Carriere (2002) are sub-optimal for the kink-estimation and would lead to asymptotic confidence intervals which are too wide as well. Lin et al. (2008) derived asymptotic normality of an exponent of their γ-th kink in a random design with dependency structure if γ is odd on the one hand, but on the other hand if γ is even their estimate converges to a non-standard limit distribution. Thus, similarly as in Müller (1992) the results of Lin et al. (2008) do not seem to be optimal for the construction of confidence sets of kink-locations of higher order.

The only work which addresses kink-location estimation of first order based on asymptotic normality without any drawbacks seems to be Eubank and Speckman (1994). Recently, Mallik et al. (2013) constructed conservative asymptotic confidence intervals for the kink-location in a nonparametric regression model with equidistant design for a specific shape of the regression function by a p-value based method. Their assumptions on the smoothness of the regression function outside the kink is milder than those made in the aforementioned literature on kink estimation respectively related problems of it. Nevertheless, their shape conditions are rather strict.

Chapter 3 of this thesis, which is based on Bengs and Holzmann (2018), provides a rate optimal method to construct confidence intervals for the single kink-location of order γ of regression functions, which are assumed to have at least $s \geq \gamma + 1$ continuous derivatives away from the kink-location and without requiring any specific shape conditions. Furthermore, based on a Lepski-choice of the bandwidth, the resulting confidence intervals are adaptive with respect to s over smaller, separated function classes which allow for an explicit control of the bias term.

CHAPTER 2

Asymptotic confidence sets for the jump curve in bivariate regression problems

In this chapter, we construct asymptotic confidence bands for the single edge in an otherwise smooth image function based on the rotational difference kernel method by Qiu (1997) or Müller and Song (1994). Using methods from M-estimation, we show consistency of the estimators of location and slope of the edge function and develop a linearization of the contrast process which is uniform in this bivariate parameter. The uniform confidence bands then rely on a Gaussian approximation of the score process together with anti-concentration results for suprema of Gaussian processes from Chernozhukov et al. (2014), while pointwise bands are based on asymptotic normality. A technical difficulty in the problem are the distinct rates of the estimators of location and slope, which will be coped with a reparametrization of the criterion function.

The chapter is structured as follows. In Section 2.1, we recall the model as well as the estimators for location and slope of the edge. The main theoretical results can be found in Section 2.2, while Section 2.3 contains a simulation study and an illustrative application of the proposed method to a real-world image. An outline of the proofs of the main results is provided in Section 2.4, while full technical details are deferred to the Section 2.6. Finally, Section 2.5 discusses some extensions and modifications of the results.

2.1 Model and estimate

We consider a bivariate boundary fragment (see (1.12)) of a noisy gray scale image, in which real random variables Y_{i_1,i_2} are observed according to the model

$$Y_{i_1,i_2} = m_\phi(\mathbf{x_{i_1,i_2}}) + \varepsilon_{i_1,i_2}, \quad (i_1,i_2) \in \{1,\dots,n\}^2, \tag{2.1}$$

where $\mathbf{x_{i_1,i_2}} = (x_{i_1}, x_{i_2})^T$ form a deterministic, regular rectangular grid in $[0,1]^2$, and m_ϕ is an unknown square-integrable function on $[0,1]^2$, which is sufficiently smooth besides a discontinuity curve ϕ. Specifically, we assume that m_ϕ is of the form

$$\begin{aligned} m_\phi(x,y) &= m(x,y) + j_{\tau,\phi}(x,y), \\ j_{\tau,\phi}(x,y) &= j_\tau(x,y) = \tau(x) 1_{[0,\phi(x)]}(y), \end{aligned} \tag{2.2}$$

where $m : [0,1]^2 \to \mathbb{R}$ is the smooth part of the image, $\tau : [0,1] \to \mathbb{R}_+$ the jump-height-curve and $\phi : [0,1] \to (0,1)$ the jump-location-curve.

The used estimation procedure for this chapter is the rotated difference kernel as described by equations (1.13) – (1.16), that is for $h > 0$, $\mathbf{z} = (z_1, z_2)^T \in [0,1]^2$ and $\psi \in [-\pi/2, \pi/2]$ the considered criterion function is

$$\hat{\mathrm{M}}_n(\mathbf{z}; \psi, h) = n^{-2} \sum_{i_1, i_2 = 1}^{n} Y_{i_1, i_2} K(\mathbf{z} - \mathbf{x}_{\mathbf{i_1}, \mathbf{i_2}}; \psi, h),$$

where

$$K(\mathbf{z}; \psi, h) = K(h^{-1} \mathbf{D}_{-\psi} \mathbf{z})/h^2, \quad \mathbf{D}_\psi = \begin{pmatrix} \cos(\psi) & \sin(-\psi) \\ \sin(\psi) & \cos(\psi) \end{pmatrix},$$

as well as $K(\mathbf{z}) = K(z_1, z_2) = K_1(z_1) K_2(z_2)$. For our asymptotic analysis, we require the following set of assumptions.

Assumption 2.1 (Errors). The errors $\boldsymbol{\varepsilon}_{i_1, i_2}$ are square-integrable, centered, independent and identically distributed random variables with standard deviation $\sigma > 0$. Moreover, $\mathbb{E}|\varepsilon_{1,1}|^5 < \infty$. ◊

Assumption 2.2 (Smoothness). We have that $\phi \in C^2[0,1], \tau \in C^2(\mathbb{R}_+)$ and $m \in C^2[0,1]^2$. ◊

Assumption 2.3 (Kernel). The kernel functions K_1 and K_2 are three-times continuously differentiable with support in $[-1,1]$ and satisfy the following conditions.

1. K_1 is symmetric, i.e. $K_1(x) = K_1(-x)$, and $K_1(x) > 0$ for $x \in (-1,1)$. Further, K_1 satisfies

$$\int_{[-1,1]} K_1(x) \, \mathrm{d}x = 1 \quad \text{and} \quad K_1^{(j)}(-1) = K_1^{(j)}(1) = 0, \; j = 0,1,2.$$

2. K_2 is an odd function, i.e. $K_2(x) = -K_2(-x)$, in particular $K_2(0) = 0$, and satisfies $\int_0^1 x K_2(x) \, \mathrm{d}x = 0$ as well as

$$\int_0^1 K_2(x) \, \mathrm{d}x = -\int_{-1}^0 K_2(x) \, \mathrm{d}x = 1, \; K_2^{(1)}(0) > 0 \quad \text{and} \quad K_2^{(j)}(-1) = K_2^{(j)}(1) = 0, \quad j = 0,1,2.$$

◊

Assumption 2.4 (Bandwidth). The bandwidth $h = h_n$ lies in the range

$$h_n \in \left[C_1 \log(n)^\eta / n^{1/2}, C_2 / n^{1/3} \right]$$

for some finite constants $C_1, C_2 > 0$ and some fixed $\eta > 1/2$. ◊

Note that this in particular implies that $n h_n^2 \to \infty$, a standard assumption in bivariate nonparametric estimation and that $\log(n)$ and $\log(h^{-1})$ are of the same order.

2.2 Asymptotic theory

In the following, we fix some compact subinterval $I \subset (0,1)$. Let us start with uniform consistency of the estimators on I. Recall the estimates of the method which are given in (1.15) or (1.16), i.e. for any $x \in I$

$$\left(\hat{\phi}_n(x), \hat{\psi}_n(x)\right) \in \underset{y \in [h, 1-h], \psi \in [-\pi/2, \pi/2]}{\arg\max} \hat{\mathrm{M}}_n((x,y)^T; \psi, h), \quad \text{and} \quad \hat{\tau}_n(x) = \hat{\mathrm{M}}_n((x, \hat{\phi}_n(x))^T; \hat{\psi}_n(x), h).$$

Theorem 2.1. *In model (2.1), under the Assumptions 2.1 – 2.4 we have that*

$$\sup_{x\in I}\left|\hat{\phi}_n(x)-\phi(x)\right| = O_P(\log(n)^{1/2}/n), \quad \sup_{x\in I}\left|\hat{\psi}_n(x)-\psi(x)\right| = O_P(\log(n)^{1/2}/nh),$$
$$\sup_{x\in I}\left|\hat{\tau}_n(x)-\tau(x)\right| = O_P(h)+O_P(\log(n)^{1/2}/nh).$$

As expected the rate of convergence for ψ is slower than for ϕ, as estimation of ψ corresponds to estimation of a derivative. Thus, we encounter here the distinct rate issue already mentioned in Section 1.3, which will be discussed in Remark 1 below.

Now we turn to asymptotic normality of $\big(\hat{\phi}_n(x),\hat{\psi}_n(x)\big)$ for fixed x. We shall require the maximizers of the deterministic version of the contrast function

$$\big(\phi_n(x),\psi_n(x)\big) \in \operatorname*{arg\,max}_{y\in[h,1-h],\psi\in[-\pi/2,\pi/2]} \mathbb{E}\big(\hat{\mathrm{M}}_n((x,y)^T;\psi,h)\big), \tag{2.3}$$

where $\hat{\mathrm{M}}_n$ is given in (1.14) resp. in Section 2.1.

Theorem 2.2. *In model (2.1), under the Assumptions 2.1 – 2.4 we have for any $x\in I$ that*

$$(n,nh)\begin{pmatrix}\hat{\phi}_n(x)-\phi_n(x)\\ \hat{\psi}_n(x)-\psi_n(x)\end{pmatrix} \rightsquigarrow N_2\Big(\mathbf{0},\sigma^2\Sigma(x)\Big),$$

where $\Sigma(x) = diag\big(V_N(x)/V_H^2(x),\, W_N/W_H^2(x)\big)$ *with*

$$\begin{aligned}
V_H(x) &= \tau(x)\cos^2(\psi(x))K_2^{(1)}(0), \qquad W_H(x) = \tau(x)K_2^{(1)}(0)\int_{-1}^1 y^2K_1(y)\,\mathrm{d}y,\\
W_N &= \int_{-1}^1\Big(\int_{-1}^1\big(K_1(z_1)K_2^{(1)}(z_2)z_1-K_1^{(1)}(z_1)K_2(z_2)z_2\big)^2\mathrm{d}z_1\Big)\mathrm{d}z_2,\\
V_N(x) &= \sin^2(\psi(x))\int_{-1}^1\Big(\int_{-1}^1\big(K_1^{(1)}(z_1)K_2(z_2)\big)^2\mathrm{d}z_1\Big)\mathrm{d}z_2+\cos^2(\psi(x))\int_{-1}^1\Big(\int_{-1}^1\big(K_1(z_1)K_2^{(1)}(z_2)\big)^2\mathrm{d}z_1\Big)\mathrm{d}z_2.
\end{aligned} \tag{2.4}$$

Remark 1 (Rescaling of the contrast function). The asymptotic distribution in Theorem 2.2 may be derived using the techniques described in Section 1.3.2, which are however not straightforward to apply because of the distinct rates. In order to deal with them, we follow the dilated criterion function idea noted in Section 1.3.2, which is to center and rescale the y-argument of the contrast function $\hat{\mathrm{M}}_n((x,y)^T;\psi,h)$ as $y=\phi(x)+wh$.

Define for any $x\in I$ the set of the rescaled parameter for the jump-positions as

$$B_{n,x} = \big\{\, w\in\mathbb{R} : \phi(x)+wh \in \big[h,1-h\big]\,\big\},$$

and for $(x,\psi,w)^T\in I\times\mathbb{R}^2$ define the rescaled contrast function

$$\hat{\mathbb{M}}_n(w,\psi;x) = \hat{\mathrm{M}}_n((x,\phi(x)+hw)^T;\psi,h), \tag{2.5}$$

where $\hat{\mathrm{M}}_n$ is given in (1.14) resp. in Section 2.1, so that for the maximizers

$$\big(\hat{w}_n(x),\hat{\psi}_n(x)\big) \in \operatorname*{arg\,max}_{w\in B_{n,x},\psi\in[-\pi/2,\pi/2]} \hat{\mathbb{M}}_n(w,\psi;x)$$

we have that $\hat{w}_n(x) = \big(\hat{\phi}_n(x) - \phi(x)\big)/h$. Then Theorem 2.2 is equivalent to

$$nh\left[\big(\hat{w}_n(x), \hat{\psi}_n(x)\big)^T - \big(w_n(x), \psi_n(x)\big)^T\right] \rightsquigarrow N_2\Big(\mathbf{0}, \sigma^2 \Sigma(x)\Big)$$

where $w_n(x) = \big(\phi_n(x) - \phi(x)\big)/h$ and where the covariance matrix is as specified in the theorem. $\diamond$

Remark 2 (Confidence intervals). In order to construct asymptotic confidence intervals, we choose a consistent estimate $\hat{\sigma}_n^2$ of the error variance σ^2, see e.g. Munk et al. (2005) and Section 2.3. Given $\alpha \in (0,1)$ we obtain an asymptotic level $(1-\alpha)$ confidence interval $\big[c_\phi^-(x), c_\phi^+(x)\big]$ for $\phi_n(x)$ by letting

$$c_\phi^\pm(x) = \hat{\phi}_n(x) \pm \frac{\hat{\sigma}_n \hat{V}_N^{1/2}(x)\, q_{1-\alpha/2}}{n \hat{V}_H(x)}, \tag{2.6}$$

where $q_\beta = q_\beta(N_1(0,1))$ is the β-quantile of $N(0,1)$ and

$$\begin{aligned} \hat{V}_H(x) &= \hat{\tau}_n(x) \cos^2(\hat{\psi}_n(x))\, K_2^{(1)}(0), \\ \hat{V}_N(x) &= \sin\big(\hat{\psi}_n(x)\big)^2 \int_{-1}^1 \int_{-1}^1 (K_1^{(1)}(z_1) K_2(z_2))^2 \,\mathrm{d}z_1 \mathrm{d}z_2 \\ &\quad + \cos\big(\hat{\psi}_n(x)\big)^2 \int_{-1}^1 \int_{-1}^1 (K_1(z_1) K_2^{(1)}(z_2))^2 \,\mathrm{d}z_1 \mathrm{d}z_2. \end{aligned} \tag{2.7}$$

Similarly, for $\psi_n(x)$ we get an asymptotic level $(1-\alpha)$ confidence interval $\big[c_\psi^-(x), c_\psi^+(x)\big]$ by

$$c_\psi^\pm(x) = \hat{\psi}_n(x) \pm \frac{\hat{\sigma}_n W_N^{1/2} q_{1-\alpha/2}}{nh \hat{W}_H(x)}, \qquad \hat{W}_H(x) = \hat{\tau}_n(x) K_2^{(1)}(0) \int_{-1}^1 y^2 K_1(y)\,\mathrm{d}y. \tag{2.8}$$

These confidence intervals can be used for the actual parameters as well. Recall from Section 1.2 that the bias for a regular deterministic design is not negligible in the jump-curve estimation setting. Nevertheless, we will see in our simulation study that the bias often is reasonably small and the aforementioned confidence intervals have good coverage also for $(\phi(x), \psi(x))$. $\diamond$

Now let us turn to the construction of uniform confidence bands. For independent standard normally distributed random variables $\xi_{1,1}, \ldots, \xi_{n,n}$ which are independent of Y_{i_1,i_2} as well, consider the process

$$\tilde{Z}_n^\phi(x) := \frac{1}{nh\big(\hat{V}_N(x)\big)^{1/2}} \sum_{i_1,i_2=1}^n \xi_{i_1,i_2} \Big\langle (\nabla K)(h^{-1} \mathbf{D}_{-\hat{\psi}_n(x)}((x, \hat{\phi}_n(x))^T - \mathbf{x_{i_1,i_2}})), (\sin\hat{\psi}_n(x), \cos\hat{\psi}_n(x))^T \Big\rangle, \tag{2.9}$$

where $\nabla K(z_1, z_2) = (K_1^{(1)}(z_1) K_2(z_2), K_1(z_1) K_2^{(1)}(z_2))^T$ and by definition of the rotation matrix in (1.13)

$$\mathbf{D}_{-\hat{\psi}_n(x)}((x, \hat{\phi}_n(x))^T - \mathbf{x_{i_1,i_2}}) = \begin{pmatrix} \cos(\hat{\psi}_n(x))(x - x_{i_1}) + \sin(\hat{\psi}_n(x))(\hat{\phi}_n(x) - x_{i_2}) \\ \cos(\hat{\psi}_n(x))(\hat{\phi}_n(x) - x_{i_2}) - \sin(\hat{\psi}_n(x))(x - x_{i_1}) \end{pmatrix},$$

as well as

$$\tilde{Z}_n^\psi(x) := \frac{1}{nh^2\big(W_N\big)^{1/2}} \sum_{i_1,i_2=1}^n \xi_{i_1,i_2} \Big\langle (\nabla K)(h^{-1} \mathbf{D}_{-\hat{\psi}_n(x)}((x, \hat{\phi}_n(x))^T - \mathbf{x_{i_1,i_2}})), \mathbf{D}_{3\pi/2-\hat{\psi}_n(x)}((x, \hat{\phi}_n(x))^T - \mathbf{x_{i_1,i_2}}) \Big\rangle. \tag{2.10}$$

These processes correspond to the centered score processes evaluated at the estimates $(\hat{\phi}_n, \hat{\psi}_n)$ with independent noise-variables ξ_{i_1,i_2}. Furthermore, set

$$\tilde{\mathbf{M}}_\phi := \sup_{x\in I} |\tilde{Z}_n^\phi(x)|, \qquad \tilde{\mathbf{M}}_\psi := \sup_{x\in I} |\tilde{Z}_n^\psi(x)|, \tag{2.11}$$

the quantiles of which may be determined by simulations. The following result is the basis for constructing uniform confidence sets.

Theorem 2.3. *Consider model (2.1) under the Assumptions 2.1 – 2.4, and assume that $\hat{\sigma}_n$ is an estimator for σ which satisfies $P(|\hat{\sigma}_n/\sigma - 1| \geq s_n) = o(1)$ for which for some sequence $s_n = o(\log(n)^{-1})$. Then for $\alpha \in (0,1)$, it holds that $q_{1-\alpha}(\tilde{\mathbf{M}}_\phi) = O((\log n)^{1/2})$ and for any sequence $t_n = o(1)$ such that $t_n\sqrt{\log(n)} \to \infty$ we have that*

$$\liminf_n P\Big(\sup_{x\in I}\Big|\frac{\hat{V}_H(x)(\hat{\phi}_n(x)-\phi(x))}{\hat{V}_N^{1/2}(x)}\Big| \leq \frac{(1+t_n)\,\hat{\sigma}_n\, q_{1-\alpha}(\tilde{\mathbf{M}}_\phi)}{n}\Big) \geq 1-\alpha$$

as well as $q_{1-\alpha}(\tilde{\mathbf{M}}_\psi) = O((\log n)^{1/2})$ and

$$\liminf_n P\Big(\sup_{x\in I}\Big|\frac{\hat{W}_H(x)(\hat{\psi}_n(x)-\psi(x))}{W_N^{1/2}}\Big| \leq \frac{(1+t_n)\,\hat{\sigma}_n\, q_{1-\alpha}(\tilde{\mathbf{M}}_\psi)}{nh}\Big) \geq 1-\alpha.$$

Remark 3 (Asymptotic confidence band). Given $\alpha \in (0,1)$ a confidence band for ϕ which is asymptotically conservative at level $(1-\alpha)$ is given by

$$\Big\{[c^-_{\phi,u}(x), c^+_{\phi,u}(x)] \mid x \in I\Big\}, \quad c^\pm_{\phi,u}(x) = \hat{\phi}_n(x) \pm \frac{(1+t_n)\hat{\sigma}_n\,\hat{V}_N^{1/2}(x)\, q_{1-\alpha}(\tilde{\mathbf{M}}_\phi)}{n\,\hat{V}_H(x)}. \tag{2.12}$$

Similarly, an asymptotic level $(1-\alpha)$ confidence band for the jump-slope curve ψ is obtained by

$$\Big\{[c^-_{\psi,u}(x), c^+_{\psi,u}(x)] \mid x \in I\Big\}, \quad c^\pm_{\psi,u}(x) = \hat{\psi}_n(x) \pm \frac{(1+t_n)\hat{\sigma}_n\, W_N^{1/2}\, q_{1-\alpha}(\tilde{\mathbf{M}}_\psi)}{nh\,\hat{W}_H(x)}. \tag{2.13}$$

Note that these are versions of the asymptotic pointwise confidence intervals in (2.6) resp. (2.8), corrected by the logarithmic factor $q_{1-\alpha}(\tilde{\mathbf{M}}_\phi)$ for uniform coverage. Further, the uniform confidence bands also directly apply to the actual parameters $(\phi(x), \psi(x))$, the price to pay being that they are asymptotically conservative. $\diamond$

2.3 Simulations

In this section we investigate the finite sample properties of the proposed asymptotic confidence sets for the location $\phi(x)$ of the edge as well as of the estimator

$$\big(\hat{\sigma}_n^2 \cdot \hat{V}_N(x)/\hat{V}_H^2(x)\big)^{1/2} \tag{2.14}$$

for the asymptotic standard deviation of $n\hat{\phi}_n(x)$ in Theorem 2.2 using (2.7). Further, we also investigate the bias in the estimation of the edge when using a deterministic rectangular grid.

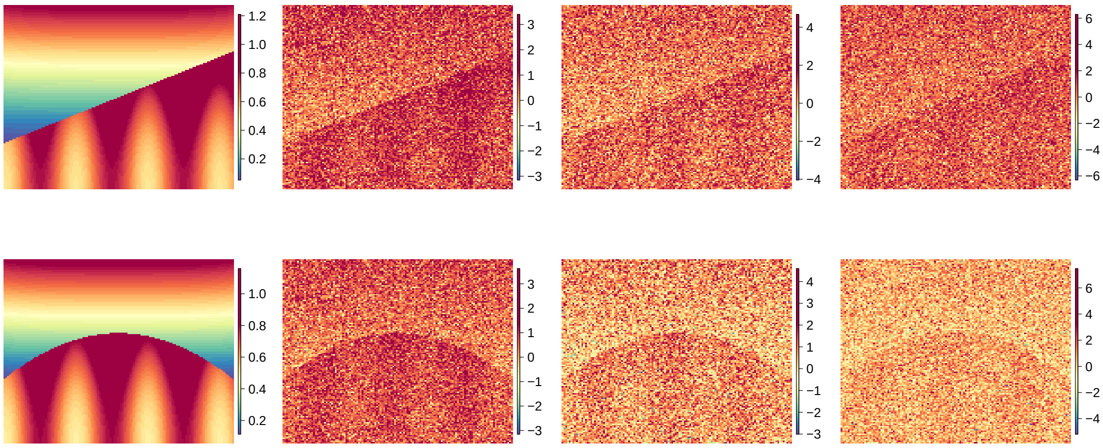

Figure 2.1.: Upper panel: Image function m_{ϕ_1} with $\tilde{\sigma} = 0,\ 0.5,\ 0.7,\ 0.9$ (from left to right) and $n = 128$ respectively.
Lower panel: Image function m_{ϕ_2} with $\tilde{\sigma} = 0,\ 0.5,\ 0.7,\ 0.9$ (from left to right) and $n = 128$ respectively.

Simulation setup

We choose

$$m(x,y) = \sin(y^2)\cos\left((x-1/2)^2\right), \quad \tau(x) = 3\sin^2(10x)/10 + 1/2,$$

as background image and jump height, and consider the following two edge functions

$$\phi_1(x) = 1/4 + x/2, \quad \phi_2(x) = -(x-1/2)^2 + 3/5,$$

from which we form the regression functions m_{ϕ_i}, $i = 1,2$, according to model (2.2).

Further, we choose $\epsilon \sim t_{10}(0,\tilde{\sigma})$ i.e. a student-t-distribution with location parameter zero, scale parameter $\tilde{\sigma}$ and ten degrees of freedom. Thus, the noise-level is given by $\sigma = \tilde{\sigma}\sqrt{10/8}$. For illustration purposes we display observations from the two models in Figure 2.1 for grid size $n^2 = 128^2$.

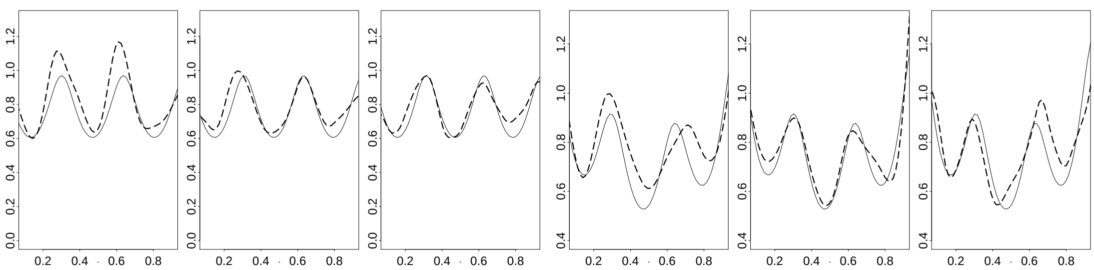

Figure 2.2.: Asymptotic standard deviation for $n\hat{\phi}_n$ (solid lines), and its estimates (dotted-dashed lines). Three leftmost pictures: Standard deviation estimation for the image-function m_{ϕ_1} for $\tilde{\sigma} = 0.5$ and $n = 128, 196, 256$ (from left to right). Three rightmost pictures: Standard deviation estimation for the image-function m_{ϕ_2} for $\tilde{\sigma} = 0.5$ and $n = 128, 196, 256$ (from left to right).

We use the kernels

$$K_1(x) = C_1 \exp\left(-(1-x^2)^{-1}\right) 1_{[-1,1]}(x), \quad K_2(x) = C_2 \exp\left(-(1-x^2)^{-1}\right) \left(x^3 - x\right) 1_{[-1,1]}(x),$$

where C_1 and C_2 are normalizing constants such that $\int K_1 = 1$ and $\int_0^1 K_2 = 1$.

Concerning the bandwidth h, one could choose the bandwidth according to a selection rule like cross-validation or Lepski's rule to optimally estimate the background image m. Qiu (2005) discusses simpler, more heuristic alternatives, one of which is to choose the window so that it contains approximately 100 design points. Although this is certainly not a universal rule which works for any n, we achieved reasonably good results with this approach for our grid sizes of $n^2 \in \{128^2, 196^2, 256^2\}$. In repeated simulations we used 500 repetitions.

Estimating the asymptotic standard deviation

We start by investigating the numerical performance of the estimator (2.14) of the asymptotic standard deviation. To this end, we need to specify $\hat{\sigma}_n$, for which we choose squares of differences of all neighboring observation pairs properly normalized. The theory in Munk et al. (2005) does not immediately apply when estimating on the full image. One possibility is to restrict estimation to a smooth part of the image. We also simulated the estimator $\hat{\sigma}_n$ of the standard deviation σ separately, the results (not displayed) were also satisfactory.

Table 2.1.: Root of the MSE of the standard deviation estimation for some points x if $\tilde{\sigma} = 0.5$ resp. $\tilde{\sigma} = 0.9$ and scenario ϕ_1. The last row indicates the mean of the RMSE for 64 points in the corresponding setting.

ϕ_1	$\tilde{\sigma} = 0.5$				$\tilde{\sigma} = 0.9$			
x	$n = 128$	$n = 196$	$n = 256$	asymp.sd	$n = 128$	$n = 196$	$n = 256$	asymp.sd
0.040	0.104	0.110	0.112	0.888	0.232	0.183	0.201	1.599
0.142	0.101	0.080	0.071	0.611	0.209	0.154	0.119	1.100
0.347	0.144	0.128	0.122	0.913	0.272	0.208	0.200	1.644
0.449	0.107	0.084	0.075	0.617	0.183	0.163	0.136	1.111
0.653	0.186	0.160	0.136	0.935	0.325	0.222	0.244	1.683
0.858	0.160	0.132	0.123	0.725	0.275	0.206	0.204	1.305
	0.148	0.124	0.111		0.264	0.195	0.176	

Table 2.2.: Root of the MSE of the standard deviation estimation for some points x if $\tilde{\sigma} = 0.5$ resp. $\tilde{\sigma} = 0.9$ and scenario ϕ_2. The last row indicates the mean of the RMSE for 64 points in the corresponding setting.

ϕ_2	$\tilde{\sigma} = 0.5$				$\tilde{\sigma} = 0.9$			
x	$n = 128$	$n = 196$	$n = 256$	asymp.sd	$n = 128$	$n = 196$	$n = 256$	asymp.sd
0.040	0.137	0.148	0.153	1.149	0.274	0.244	0.243	2.069
0.142	0.130	0.087	0.087	0.691	0.226	0.170	0.158	1.244
0.347	0.142	0.124	0.132	0.840	0.257	0.196	0.202	1.511
0.449	0.101	0.078	0.067	0.540	0.191	0.133	0.123	0.972
0.653	0.193	0.156	0.143	0.859	0.317	0.238	0.182	1.547
0.858	0.107	0.101	0.100	0.820	0.207	0.168	0.171	1.477
	0.143	0.119	0.111		0.263	0.200	0.182	

Table 2.3.: Average coverage and width of the pointwise confidence intervals for the jump-location in (2.6) for $\tilde{\sigma} = 0.5$ and $\tilde{\sigma} = 0.9$ over 64 design points.

	$\tilde{\sigma} = 0.5$				$\tilde{\sigma} = 0.9$			
	95% nominal coverage		99% nominal coverage		95% nominal coverage		99% nominal coverage	
	coverage	width	coverage	width	coverage	width	coverage	width
				$n = 128$				
ϕ_1	0.960	0.025	0.991	0.033	0.943	0.043	0.985	0.056
ϕ_2	0.958	0.025	0.991	0.033	0.946	0.042	0.983	0.056
				$n = 196$				
ϕ_1	0.958	0.016	0.992	0.021	0.944	0.028	0.990	0.037
ϕ_2	0.951	0.016	0.988	0.022	0.943	0.028	0.988	0.037
				$n = 256$				
ϕ_1	0.958	0.012	0.992	0.016	0.941	0.021	0.997	0.028
ϕ_2	0.949	0.013	0.988	0.017	0.949	0.021	0.993	0.029

Table 2.4.: Average coverage and width of the uniform confidence bands for the jump-location in (2.12) for $\tilde{\sigma} = 0.5$ and $\tilde{\sigma} = 0.9$.

	$\tilde{\sigma} = 0.5$				$\tilde{\sigma} = 0.9$			
	95% nominal coverage		99% nominal coverage		95% nominal coverage		99% nominal coverage	
	coverage	width	coverage	width	coverage	width	coverage	width
				$n = 128$				
ϕ_1	0.943	0.051	0.986	0.059	0.955	0.062	0.999	0.071
ϕ_2	0.959	0.049	0.984	0.059	0.945	0.065	0.999	0.075
				$n = 196$				
ϕ_1	0.945	0.032	0.988	0.038	0.954	0.039	0.999	0.045
ϕ_2	0.957	0.032	0.986	0.039	0.948	0.042	1.000	0.050
				$n = 256$				
ϕ_1	0.949	0.025	0.987	0.028	0.951	0.030	0.993	0.032
ϕ_2	0.953	0.025	0.994	0.033	0.955	0.032	0.999	0.039

Next we present the results for (2.14). Figure 2.2 shows smoothed estimates for specific samples for grid sizes $n^2 \in \{128^2, 196^2, 256^2\}$ for the two edge functions ϕ_i. Further, in Tables 2.1 and 2.2 we plot the square roots of the Mean-Squared-Error (RMSE) of the standard deviation estimates for the three sample sizes and two edge curves at various observation points x based on repetitions. For purposes of comparison the actual asymptotic standard deviation is given as well. One observes that the RMSE in most settings decreases as the number of grid points increases. Further, the magnitude of the RMSE as compared to the actual value of the asymptotic standard deviation is quite small for all cases.

Confidence intervals and confidence bands

We investigate the coverage behavior and average width of (2.6) as well as of (2.12) for the true jump-location-curves ϕ_i in both settings $i = 1,2$. The results are summarized in Tables 2.3 and 2.4 for the noise-levels $\tilde{\sigma} = 0.5$ and $\tilde{\sigma} = 0.9$. The values in the tables of the pointwise confidence intervals correspond to the average of the respective quantity over 64 design points x. The quantile $q_\beta(\tilde{\mathbf{M}}_\phi)$ for $\beta \in (0,1)$ was simulated based on a multiplier bootstrap sample of size 40000. Furthermore, as there is no explicit representation of the t_n-term we have chosen it as given in Table 2.5 for the different scenarios. The t_n decrease for increasing sample size n and are of the same magnitude for both scenarios, that is for ϕ_1 resp. ϕ_2. By way of comparison, we give the values of $\log(n)^{-1/2}$ for the different sample sizes n as well. Especially, in the high-noise case the magnitude of our choice for t_n is much smaller as this benchmark.

Overall, the simulated coverage probabilities for the pointwise confidence intervals are reasonably

Table 2.5.: Choice of t_n in (2.12).

	$\tilde{\sigma} = 0.5$			$\tilde{\sigma} = 0.9$		
	$n = 128$	$n = 196$	$n = 256$	$n = 128$	$n = 196$	$n = 256$
ϕ_1	0.37	0.34	0.335	0.07	0.001	0
ϕ_2	0.4	0.37	0.25	0.14	0.1	0.06
$1/\sqrt{\log(n)}$	0.45	0.44	0.42	0.45	0.44	0.42

close to their nominal values in all scenarios, and the intervals become narrower with increasing numbers of grid points. As expected from the theoretical developments, the uniform confidence bands are somewhat conservative particular in the high-noise level case. Figure 2.3 illustrates the estimated curve as well as the confidence intervals and bands for ϕ_1 and ϕ_2 in the low-noise-level for increasing numbers of grid points for $\alpha = 0.05$, that is, asymptotic 95% coverage probability. Apparently, the variability of the jump-location estimator decreases and the confidence intervals resp. bands become narrower. Besides, the confidence bands adapt to the shape of the pointwise confidence intervals as the width-terms only differ in the choice of the quantile. We omitted the results for the confidence intervals in (2.13) for $\psi(x)$, since in all settings the empirical standard deviation of the estimator $\hat{\psi}_n$ was much smaller than the estimated resp. asymptotic standard deviation, which led to conservative confidence intervals having high coverage probabilities for even moderate significance levels.

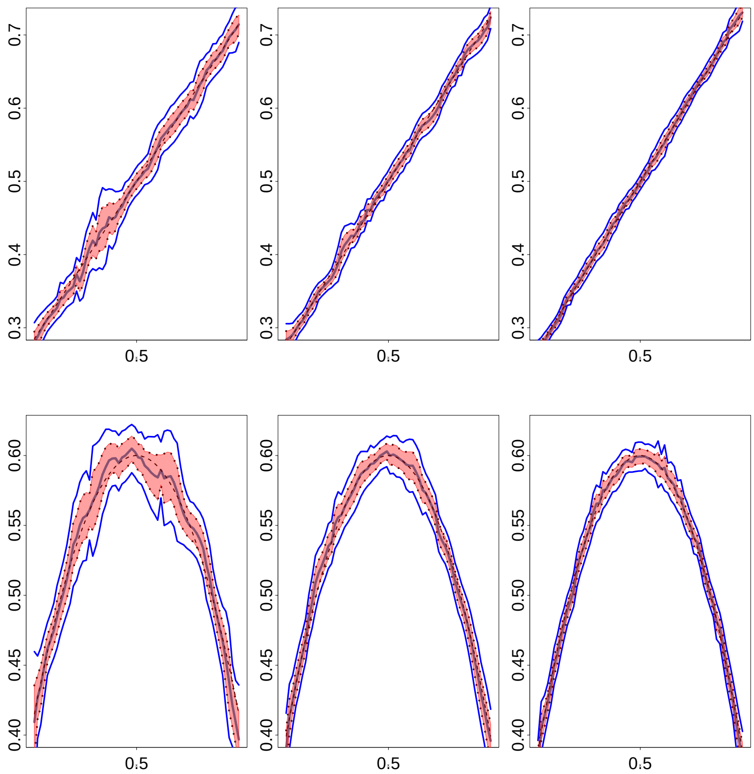

Figure 2.3.: Top panel: 95% Confidence intervals and estimate of the jump-location-curve (shaded area and solid line within), uniform confidence bands (solid lines) and true jump-location-curve ϕ_1 (dashed lines inside shaded area) for $n = 128, 196, 256$ and $\tilde{\sigma} = 0.5$. Lower panel: 95% Confidence intervals and estimate of the jump-location-curve (shaded area and solid line within), uniform confidence bands (solid lines) and true jump-location-curve ϕ_2 (dashed lines inside shaded area) for $n = 128, 196, 256$ and $\tilde{\sigma} = 0.5$.

Comparing bias and standard deviation

The previous results for the coverage of the pointwise confidence intervals for the true jump-location are somewhat surprising since it is well-known that jump-curve estimation in a nonparametric regression setting with fixed design can have an asymptotic bias of order $O(n^{-1})$.
Therefore, we also investigated the order of the bias numerically and compared it to the standard deviation. Tables 2.7 and 2.6 present the results for the ratio of the bias and the standard deviation for different design points x in the low-noise- and high-noise-level-case. The ratios are quite small, showing that the bias indeed is often of quite smaller magnitude than the standard deviation even in case of a fixed design.

Table 2.6.: Ratio between computed bias and estimated standard deviation for different points x in scenario ϕ_1. The last line contains the average ratio over 64 design points for $\tilde{\sigma} = 0.5$ and $\tilde{\sigma} = 0.9$.

ϕ_2	$\tilde{\sigma} = 0.5$			$\tilde{\sigma} = 0.9$		
x	$n = 128$	$n = 196$	$n = 256$	$n = 128$	$n = 196$	$n = 256$
0.040	0.091	0.121	0.051	0.037	0.021	0.042
0.142	0.227	0.267	0.172	0.059	0.116	0.096
0.347	0.181	0.216	0.030	0.041	0.144	0.073
0.449	0.091	0.337	0.189	0.061	0.305	0.063
0.653	0.120	0.095	0.079	0.020	0.081	0.117
0.858	0.218	0.223	0.161	0.093	0.149	0.116
	0.160	0.217	0.138	0.075	0.130	0.101

Table 2.7.: Ratio between computed bias and estimated standard deviation for different points x in scenario ϕ_1. The last line contains the average ratio over 64 design points for $\tilde{\sigma} = 0.5$ and $\tilde{\sigma} = 0.9$.

ϕ_1	$\tilde{\sigma} = 0.5$			$\tilde{\sigma} = 0.9$		
x	$n = 128$	$n = 196$	$n = 256$	$n = 128$	$n = 196$	$n = 256$
0.040	0.077	0.036	0.009	0.026	0.004	0.048
0.142	0.001	0.037	0.013	0.032	0.035	0.010
0.347	0.064	0.037	0.041	0.029	0.035	0.086
0.449	0.043	0.005	0.013	0.014	0.093	0.085
0.653	0.048	0.072	0.032	0.063	0.033	0.002
0.858	0.018	0.004	0.022	0.003	0.046	0.075
	0.033	0.034	0.028	0.029	0.050	0.049

Real-life image processing

Finally we apply our method to two 300x128 Grey-scale real-life images, taken by camera. They contain the outline of a rock in front of a gray background. Once, an appropriate ISO–configuration and focus on the rock and once inappropriate ISO-configuration of the camera and no focus at all are employed. In both cases we apply our method to estimate the boundary of the rock and construct 0.95-level uniform confidence sets using $t_n = 0$ and a naive estimator for the noise-level. Figure 2.4 contains the results. The jump-location-curve lies mostly inside the constructed confidence band and the width is quite satisfying, although the noise-level of the picture is rather low.

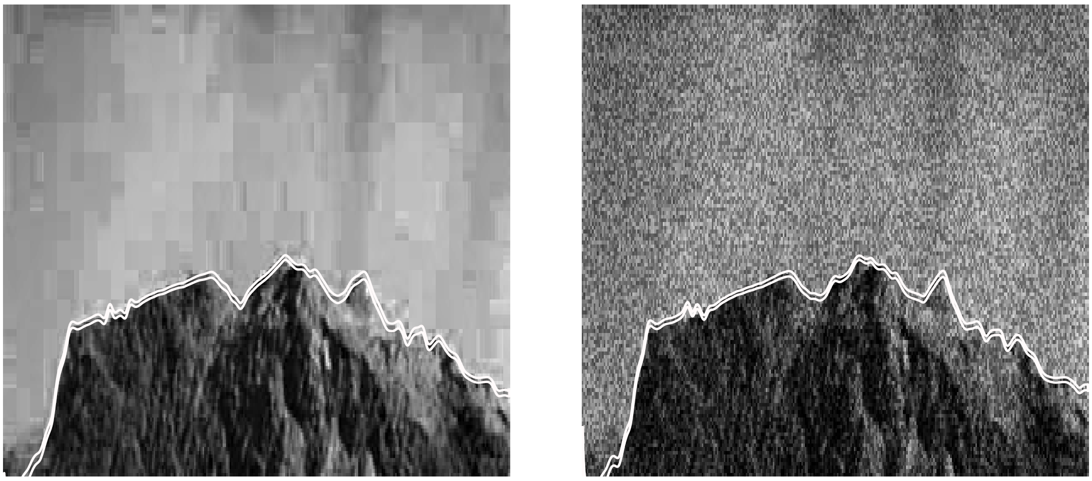

Figure 2.4.: Left: 300x128 Grey-scale picture of a rock taken from a camera with reasonable ISO-configuration and with focus on the rock. Solid lines correspond to the 95%-level confidence band for the noisy picture on the right.
Right: 300x128 Grey-scale picture of a rock taken from a camera with inappropriate ISO-configuration and with no focus. Solid lines correspond to the 95%-level confidence band for this noisy picture.

2.4 Outline of proofs

In the following we will give an outline of the proofs for the results in Section 2.2. Most of the auxiliary results used in this outline will be proved in the proceeding sections, see the remarks after each auxiliary statement.

For sake of brevity we set $\mathbf{p}(x) = \big(x, \phi(x)\big)^T$ and with $\mathbf{e_2} = (0,1)^T$ obtain that (2.5) is equivalent to

$$\hat{\mathbb{M}}_n(w,\psi;x) = \hat{\mathrm{M}}_n(\mathbf{p}(x) + wh\mathbf{e_2}; \psi, h).$$

For the deterministic maximizer we introduce the deterministic contrast function

$$\mathbb{M}_n(w,\psi;x) = \mathbb{E}\big(\hat{\mathbb{M}}_n(w,\psi;x)\big). \tag{2.15}$$

Finally, we let $\Theta_n = \bigcup_{x\in I}\{x\}\times\tilde{\Theta}_{n,x}$, where

$$\tilde{\Theta}_{n,x} = \big\{ w \in \mathbb{R} : \phi(x) + wh \in \big[h, 1-h\big] \big\} \times [-\pi/2, \pi/2].$$

From now on, we shall always assume that h is so small (n is sufficiently large) that $I \subset [h, 1-h]$ and the supremum norm $||\cdot||_\infty := ||\cdot||_{\infty,I}$ is throughout taken over I. In addition, the range of n and h values for which all occurring uniform O-terms are valid (recall the notation section at the beginning of this thesis) are implicitly given by Assumption 2.4 in this chapter.

2.4.1 Uniform consistency

In this section we show uniform consistency over I for the maximizers

$$\left(\hat{w}_n(x), \hat{\psi}_n(x)\right) \in \underset{(w,\psi)^T \in \tilde{\Theta}_{n,x}}{\arg\max} \quad \hat{\mathbb{M}}_n(w,\psi;x), \qquad x \in I,$$

where we have that $\hat{w}_n(x) = (\hat{\phi}_n(x) - \phi(x))/h$, see for instance Remark 1. The following proposition shows uniform consistency for the estimates.

Proposition 2.4. *Under the Assumptions 2.1 – 2.4 we have that*

$$|| \left(\hat{w}_n(\cdot), (\hat{\psi}_n - \psi)(\cdot)\right)^T ||_\infty \xrightarrow{P} 0.$$

Moreover,

$$|| \hat{\tau}_n - \tau ||_\infty = O(|| \hat{w}_n ||_\infty) + O(|| \hat{\psi}_n - \psi ||_\infty) + O_P(h) + O((nh)^{-1}).$$

In order to obtain Proposition 2.4, in the next two lemmas we check that the requirements of Proposition 1.2 are satisfied. The next lemma determines the limit function for $\hat{\mathbb{M}}_n(w,\psi;x)$ and shows that they fulfill the assumption (1.22).

Lemma 2.5. *Under the Assumptions 2.1 – 2.4 we have that*

$$\sup_{(x,w,\psi)\in\Theta_n} |\hat{\mathbb{M}}_n(w,\psi;x) - \mathbb{M}(w,\psi;x)| = O_P(h) + O((nh)^{-1}).$$

where

$$\mathbb{M}(w,\psi;x) = \tau(x) \int_{\mathbb{H}(\psi(x)-\psi)+w(\sin\psi,\cos\psi)^T} K(\mathbf{z})\,\mathrm{d}\mathbf{z}, \tag{2.16}$$

and

$$\mathbb{H}(\psi) = \mathbf{D}_\psi\left(\mathbb{R}\times[0,\infty)\right), \quad \psi \in [-\pi,\pi]. \tag{2.17}$$

Proof of Lemma 2.5. Split $\hat{\mathbb{M}}_n(w,\psi;x)$ into three terms,

$$\begin{aligned}
\hat{\mathbb{M}}_n(w,\psi;x) = (nh)^{-2} \sum_{i_1,i_2=1}^{n} m\left(\mathbf{x_{i_1,i_2}}\right) K\left(h^{-1}\mathbf{D}_{-\psi}\left(\mathbf{p}(x) + hw\mathbf{e_2} - \mathbf{x_{i_1,i_2}}\right)\right) \\
+ (nh)^{-2} \sum_{i_1,i_2=1}^{n} j_\tau\left(\mathbf{x_{i_1,i_2}}\right) K\left(h^{-1}\mathbf{D}_{-\psi}\left(\mathbf{p}(x) + hw\mathbf{e_2} - \mathbf{x_{i_1,i_2}}\right)\right) \\
+ (nh)^{-2} \sum_{i_1,i_2=1}^{n} \varepsilon_{i_1,i_2} K\left(h^{-1}\mathbf{D}_{-\psi}\left(\mathbf{p}(x) + hw\mathbf{e_2} - \mathbf{x_{i_1,i_2}}\right)\right) \\
=: S_n(w,\psi;x) + J_n(w,\psi;x) + E_n(w,\psi;x).
\end{aligned}$$

By taking

$$g_1(\mathbf{z}) = m(\mathbf{z}), \quad f(\mathbf{z}) = 4^{-1}\begin{pmatrix} K(\mathbf{z}) & K(\mathbf{z}) \\ K(\mathbf{z}) & K(\mathbf{z}) \end{pmatrix}, \quad g_2(\mathbf{z}) = g_3(\mathbf{z}) = (1,1)^T,$$

Lemma B.1 in the appendix with $r_1 = 2$, $r_2 = r_3 = 0$ and $j = 1$ states that

$$S_n(w,\psi;x) = m(\mathbf{p}(x) + hw\mathbf{e}_2) \int_{[-1,1]^2} K(\mathbf{z})\,\mathrm{d}\mathbf{z} + O(h) + O\left((nh)^{-1}\right) = O(h) + O\left((nh)^{-1}\right),$$

uniformly for x, w and ψ. With the same functions f, g_2, g_3 as above one obtains from Lemma B.2 that

$$J_n(w,\psi;x) = \tau(x) \int_{\mathbb{H}(\psi(x)-\psi)+w(\sin(\psi),\cos(\psi))^T} K(\mathbf{z})\,\mathrm{d}\mathbf{z} + O(h) + O\left((nh)^{-1}\right)$$
$$= \mathbb{M}(w,\psi;x) + O(h) + O\left((nh)^{-1}\right),$$

uniformly for x, w and ψ. Finally, Lemma B.3 with the same f, g_2, g_3 implies $E_n(w,\psi;x) = o_P(h)$ uniformly for x, w and ψ. This concludes the proof of the lemma. □

In the next lemma, we rewrite the asymptotic form of the contrast and show that it has a unique well-separated maximum, that is the requirement (1.23).

Lemma 2.6. *Under the Assumptions 2.2 and 2.3, it holds that if $\psi(x) - \psi = \pm\pi/2$ then $\mathbb{M}(w,\psi;x) = 0$ for all w, while otherwise*

$$\mathbb{M}(w,\psi;x) = -\tau(x) \int_{-1}^{1} K_1(y)\,\bar{K}_2(a_x(\psi)\,y + b_x(w,\psi))\,\mathrm{d}y, \qquad (2.18)$$

where

$$\bar{K}_2(y) = \int_{-1}^{y} K_2(t)\,\mathrm{d}t, \qquad a_x(\psi) = \tan(\psi(x)-\psi), \qquad b_x(w,\psi) = w\,\frac{\cos(\psi(x))}{\cos(\psi(x)-\psi)}.$$

This implies that for any $\epsilon > 0$,

$$\inf_{x\in[0,1]} \left(\mathbb{M}(0,\psi(x);x) - \sup_{(w,\psi)\in\mathbb{R}^2:\ \epsilon\le\max(|w|,|\psi(x)-\psi|)} \mathbb{M}(w,\psi;x)\right) > 0. \qquad (2.19)$$

Moreover, $\mathbb{M}(0,\psi(x);x) = \tau(x)$.

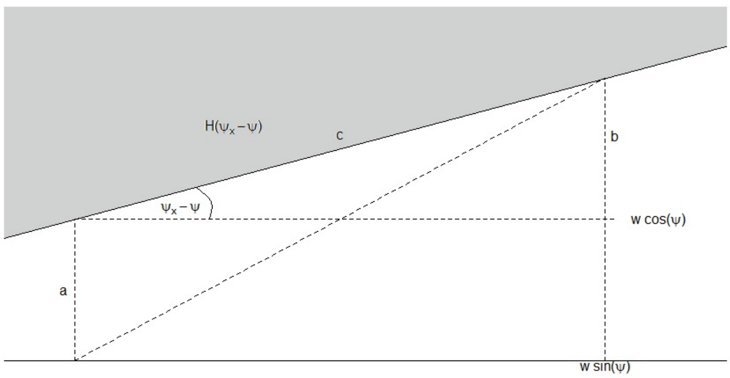

Figure 2.5.: Grey area corresponds to $\mathbb{H}(\psi(x)-\psi)$

Proof of Lemma 2.6. If $\psi(x)-\psi = \pi/2$ then $\mathbb{H}(\psi(x)-\psi) = (-\infty,0]\times\mathbb{R}$. Since $\int_{\mathbb{R}} K_2 = \int_{-1}^{1} K_2 = 0$, the statement $\mathbb{M}(w,\psi;x) = 0$ follows from (2.16), and similarly for $\psi(x)-\psi = -\pi/2$. If $\psi(x)-\psi \neq \pm\pi/2$, we start by showing that

$$\mathbb{H}(\psi(x)-\psi)+w(\sin\psi,\cos\psi)^T=\mathbb{H}(\psi(x)-\psi)+w\frac{\cos(\psi(x))}{\cos(\psi(x)-\psi)}\mathbf{e_2},$$

for an illustration see Figure 2.5. If $w=0$ the assertion is trivial. If $w\neq 0$, we need to determine a such that the vector $a\,\mathbf{e_2}$ is on the boundary of the set $\mathbb{H}(\psi(x)-\psi)+w(\sin\psi,\cos\psi)^T$, see Figure 2.5. Setting $c=w\sin(\psi)/\cos(\psi(x)-\psi)$ and $b=c\sin(\psi(x)-\psi)$ we get that, see Figure 2.5,

$$a=w\cos(\psi)-b=\frac{\cos(\psi)\cos(\psi(x)-\psi)-\sin(\psi)\sin(\psi(x)-\psi)}{\cos(\psi(x)-\psi)}=\frac{\cos(\psi(x))}{\cos(\psi(x)-\psi)}.$$

From (2.16) we obtain that

$$\begin{aligned}\mathbb{M}(w,\psi;x)&=\tau(x)\int_{\mathbb{H}(\psi(x)-\psi)+w(\sin\psi,\cos\psi)^T}K(\mathbf{z})\,\mathrm{d}\mathbf{z}\\&=\tau(x)\int_{\mathbb{H}(\psi(x)-\psi)+b_x(w,\psi)\mathbf{e}_2}K_1(z_1)K_2(z_2)\,\mathrm{d}z_1\mathrm{d}z_2\\&=\tau(x)\int_{-1}^{1}K_1(z_1)\int_{a_x(\psi)z_1+b_x(w,\psi)}^{1}K_2(z_2)\,\mathrm{d}z_2\mathrm{d}z_1\\&=\tau(x)\int_{-1}^{1}K_1(z_1)\big(\bar{K}_2(1)-\bar{K}_2(a_x(\psi)z_1+b_x(w,\psi))\big)\,\mathrm{d}z_1\\&=-\tau(x)\int_{-1}^{1}K_1(z_1)\bar{K}_2(a_x(\psi)z_1+b_x(w,\psi))\,\mathrm{d}z_1.\end{aligned}$$

Since

$$|a_x(\psi)y+b_x(w,\psi)|\geq 1\quad\text{if}\quad |w|\geq 1/|\cos(\psi(x))|+1$$

for all $y\in[-1,1]$ and $\psi\in[-\pi/2,\pi/2]$, $\mathbb{M}(w,\psi;x)$ vanishes outside a compact set of values of w and x.

Turning to (2.19), if we show that it holds for individual x, then since the supremum in (2.19) can be taken over a single compact set, we have that the left-hand side of (2.19) is a continuous function in x which is positive for any $x\in[0,1]$. Hence, the infimum over $x\in[0,1]$ is still positive.

To show that (2.19) holds for individual x, we observe that $-\bar{K}_2(0)=1$, $-\bar{K}_2(y)<1$ if $y\neq 0$, and $\bar{K}_2(y)=0$ if $|y|\geq 1$ by Assumption 2.3. Since $\mathbb{M}(w,\psi;x)$ is continuous in w and ψ, as can also be seen from (2.16), it is enough to show that $(0,\psi(x))$ is the unique maximizer of $\mathbb{M}(w,\psi;x)$. But this is immediate from (2.18) and the above properties of $\bar{K}_2$ and the positivity of K_1, since if $(w,\psi)\neq(0,\psi(x))$, there is at most a single value of z_1 for which $a_x(\psi)z_1+b_x(w,\psi)=0$ and hence $-\bar{K}_2(a_x(\psi)z_1+b_x(w,\psi))=1$. □

Proof of Proposition 2.4. The uniform consistency of the estimates $\hat{w}_n$ and $\hat{\psi}_n$ are immediate from Proposition 1.2, as Lemma 2.5 and Lemma 2.6 provide the necessary assumptions. Note that $\tau(x)=\mathbb{M}(0,\psi(x);x)$ and $\hat{\tau}_n(x)=\hat{\mathbb{M}}_n(\hat{w}_n(x),\hat{\psi}_n(x);x)$ such that

$$|\hat{\tau}_n(x)-\tau(x)|\leq|\hat{\mathbb{M}}_n(\hat{w}_n(x),\hat{\psi}_n(x);x)-\mathbb{M}(\hat{w}_n(x),\hat{\psi}_n(x);x)|+|\mathbb{M}(\hat{w}_n(x),\hat{\psi}_n(x);x)-\mathbb{M}(0,\psi(x);x)|.$$

Application of Lemma 2.5 yields

$$\sup_{x\in I}|\hat{\mathbb{M}}_n(\hat{w}_n(x),\hat{\psi}_n(x);x)-\mathbb{M}(\hat{w}_n(x),\hat{\psi}_n(x);x)|=O_P(h)+O((nh)^{-1}).$$

By the representation of $\mathbb{M}$ in (2.18) and using that

$$\partial_w\ b_x(w,\psi) = {}^{\cos(\psi(x))}\!/\!_{\cos(\psi(x)-\psi)},$$
$$\partial_\psi\ (a_x(\psi)t + b_x(w,\psi)) = -{}^{t}\!/\!_{\cos^2(\psi(x)-\psi)} - {}^{w\cos(\psi(x))\sin(\psi(x)-\psi)}\!/\!_{\cos^2(\psi(x)-\psi)},$$

we obtain by differentiation under the integral that

$$\partial_w \mathbb{M}(w,\psi;x) = -\tau(x)\int_{-1}^{1}\frac{\cos(\psi(x))}{\cos(\psi(x)-\psi)}K_1(y)K_2(a_x(\psi)y + b_x(w,\psi))\,\mathrm{d}y,$$
$$\partial_\psi \mathbb{M}(w,\psi;x) = \tau(x)\int_{-1}^{1}\left(\frac{y + w\cos(\psi(x))\sin(\psi(x)-\psi)}{\cos^2(\psi(x)-\psi)}\right)K_1(y)K_2(a_x(\psi)y + b_x(w,\psi))\,\mathrm{d}y.$$

By means of the mean value theorem,

$$|\mathbb{M}(\hat{w}_n(x),\hat{\psi}_n(x);x) - \mathbb{M}(0,\psi(x);x)| \le ||\,\hat{w}_n(x),\hat{\psi}_n(x) - \psi(x)\,||_2\ ||\,\nabla\mathbb{M}(\tilde{w},\tilde{\psi};x)\,||$$

for some $\tilde{w}$ between zero and $\hat{w}_n(x)$ and $\tilde{\psi}$ between $\psi(x)$ and $\hat{\psi}_n(x)$. Note that the components of $\nabla\mathbb{M}(\tilde{w},\tilde{\psi};x)$ are bounded as I is compact and $\tilde{w}$ is between zero and $\hat{w}_n(x)$ resp. $\tilde{\psi}$ is between $\psi(x)$ and $\hat{\psi}_n(x)$. Thus,

$$\sup_{x\in I}|\mathbb{M}(\hat{w}_n(x),\hat{\psi}_n(x);x) - \mathbb{M}(0,\psi(x);x)| = O(||\,\hat{w}_n\,||_\infty) + O(||\,\hat{\psi}_n - \psi\,||_\infty).$$

This implies that

$$||\,\hat{\tau}_n - \tau\,||_\infty = O(||\,\hat{w}_n\,||_\infty) + O(||\,\hat{\psi}_n - \psi\,||_\infty) + O_P(h) + O((nh)^{-1}),$$

as asserted. □

2.4.2 Rate of convergence: proof of Theorem 2.1

To prove the theorems in Section 2.2 we start with a simple linearization similar as (1.24) in Section 1.3.2. By the mean value theorem, we have that

$$\begin{aligned}\nabla\hat{\mathbb{M}}_n(0,\psi(x);x) &= \nabla\hat{\mathbb{M}}_n(0,\psi(x);x) - \nabla\hat{\mathbb{M}}_n\left(\hat{w}_n(x),\hat{\psi}_n(x);x\right)\\ &= -\int_0^1 \nabla\nabla^T\hat{\mathbb{M}}_n\left(t\hat{w}_n(x),\psi(x) + t\left(\hat{\psi}_n(x) - \psi(x)\right);x\right)\mathrm{d}t\ \left(\hat{w}_n(x),\hat{\psi}_n(x) - \psi(x)\right)^T,\end{aligned}$$

since $\nabla\hat{\mathbb{M}}_n\left(\hat{w}_n(x),\hat{\psi}_n(x);x\right) = 0$. This implies

$$\left(\hat{w}_n(x),\hat{\psi}_n(x) - \psi(x)\right)^T = -\hat{\mathbf{H}}_n^{-1}(x)\ \nabla\hat{\mathbb{M}}_n(0,\psi(x);x), \tag{2.20}$$

where

$$\hat{\mathbf{H}}_n(x) = \int_0^1 \nabla\nabla^T\hat{\mathbb{M}}_n\left(t\hat{w}_n(x),\psi(x) + t\left(\hat{\psi}_n(x) - \psi(x)\right);x\right)\mathrm{d}t,$$

and the existence of the inverse of $\hat{\mathbf{H}}_n(x)$ uniformly in $x \in I$ for large n and with high probability follows from Lemma 2.8 below.

Asymptotic bias

Lemma 2.7. *Under the Assumptions 2.2 – 2.4 we have that*

$$\sup_{x\in I} || nh\mathbb{E}\big(\nabla\hat{\mathbb{M}}_n(0,\psi(x);x)\big) || = O(1).$$

The proof of this lemma is provided in Section 2.6.1.

Convergence of the Hessian matrix

Lemma 2.8. *Under the Assumptions 2.1 – 2.4 it holds that*

$$||\,\hat{\mathbf{H}}_n - \mathbf{H}\,||_\infty \xrightarrow{P} 0,$$

where $\mathbf{H}(x) = diag(V_H(x), W_H(x))$ *and* $V_H(x)$ *and* $W_H(x)$ *are as in Theorem 2.2.*

The proof of Lemma 2.8 is in Section 2.6.2. Note that the limit matrix $\mathbf{H}$ corresponds to the Hessian matrix of the asymptotic criterion function $\mathbb{M}$ at the parameters $w = 0$ and $\psi = \psi(x)$. Furthermore, we only need the pointwise convergence in Lemma 2.8 to derive the proofs of Theorem 2.1 and Theorem 2.2. However, we note on advance that for the proof of Theorem 2.3 we need the uniform convergence even with an explicit rate of convergence (see Lemma 2.12).

Proposition 2.9. *Under the Assumptions 2.1 – 2.4 we have that*

$$\sup_{x\in I}||\,n\,h\,\big(\nabla\hat{\mathbb{M}}_n(0,\psi(x);x) - \mathbb{E}\big(\nabla\hat{\mathbb{M}}_n(0,\psi(x);x)\big)\,|| = O_P\big((\log n)^{1/2}\big).$$

The proof of this proposition is provided in Section 2.6.4.

Proof of Theorem 2.1. By Lemma 2.7 and Lemma 2.8

$$\sup_{x\in I}||\,\hat{\mathbf{H}}_n^{-1}(x)\;n\,h\;\mathbb{E}\big(\nabla\hat{\mathbb{M}}_n(0,\psi(x);x)\big)\,|| = O_P(1).$$

In particular, $\hat{\mathbf{H}}_n^{-1}$ is almost surely a stochastically bounded matrix-valued sequence uniformly in x. With this and Proposition 2.9,

$$\begin{aligned}\sup_{x\in I}||\,\hat{\mathbf{H}}_n^{-1}(x) n\,h\,\Big(\nabla\hat{\mathbb{M}}_n(0,\psi(x);x) - \mathbb{E}\big(\nabla\hat{\mathbb{M}}_n(0,\psi(x);x)\big)\Big) + \hat{\mathbf{H}}_n^{-1}(x)\;n\,h\;\mathbb{E}\big(\nabla\hat{\mathbb{M}}_n(0,\psi(x);x)\big)\,||\\ = O_P\big((\log n)^{1/2}\big) + O_P(1).\end{aligned}$$

From (2.20) it immediately follows that

$$||\,\big(\hat{w}_n,(\hat{\psi}_n - \psi)\big)^T\,||_\infty = O_P\big({}^{(\log n)^{1/2}}/_{nh}\big).$$

The uniform rate of convergence for $\hat{\tau}_n$ now follows by Proposition 2.4.

□

2.4.3 Asymptotic normality: proof of Theorem 2.2

Similarly as in (2.20), for the rescaled maximizers of the deterministic contrast function, that is

$$\big(w_n(x), \psi_n(x)\big) \in \underset{(w,\psi)^T \in \tilde{\Theta}_{n,x}}{\arg\max} \ \mathbb{M}_n(w, \psi; x),$$

where $\mathbb{M}_n$ as in (2.15), we have that

$$(w_n(x), \psi_n(x) - \psi(x))^T = -\mathbf{H}_n^{-1}(x) \ \nabla \mathbb{M}_n(0, \psi(x); x), \tag{2.21}$$

where $w_n(x) = \big(\phi_n(x) - \phi(x)\big)/h$ with ϕ_n as in (2.3) and

$$\mathbf{H}_n(x) = \int_0^1 \nabla \nabla^T \mathbb{M}_n\left(t\, w_n(x), \psi(x) + t\,(\psi_n(x) - \psi(x)); x\right) \mathrm{d}t.$$

Asymptotic normality of the score

Lemma 2.10. *Under the Assumptions 2.1, 2.3 and 2.4 we have that for any $x \in I$*

$$nh \Big(\nabla \hat{\mathbb{M}}_n(0, \psi(x); x) - \nabla \mathbb{M}_n(0, \psi(x); x) \Big) \rightsquigarrow N_2\Big(\mathbf{0}, \sigma^2\, diag(V_N(x), W_N)\Big).$$

Furthermore, $\nabla \mathbb{M}_n(w, \psi; x) = \mathbb{E}\big(\nabla \hat{\mathbb{M}}_n(w, \psi; x)\big)$.

Section 2.6.3 is devoted for the proof of this lemma.

In the same manner as Lemma 2.8 we have the following result for the Hessian matrix of the deterministic contrast function.

Lemma 2.11. *Under the Assumptions 2.2 – 2.4 we have that*

$$||\,\mathbf{H}_n - \mathbf{H}\,||_\infty \to 0,$$

where $\mathbf{H}(x) = diag(V_H(x), W_H(x))$ and $V_H(x)$ and $W_H(x)$ are as in Theorem 2.2.

The proof is given as well in Section 2.6.2.

Proof of Theorem 2.2. From (2.20) and (2.21), we obtain that

$$\begin{aligned} nh\, \big(\hat{w}_n(x) - w_n(x), \hat{\psi}_n(x) - \psi_n(x)\big)^T &= -\hat{\mathbf{H}}_n^{-1}(x)\ nh \Big(\nabla \hat{\mathbb{M}}_n(0, \psi(x); x) - \nabla \mathbb{M}_n(0, \psi(x); x) \Big) \\ &\quad + nh\, \nabla \mathbb{M}_n(0, \psi(x); x) \Big(\mathbf{H}_n^{-1}(x) - \hat{\mathbf{H}}_n^{-1}(x) \Big). \end{aligned}$$

On the one hand, the first term is asymptotically normally distributed with covariance matrix Σ as in the assumption by Lemmas 2.8 and 2.10 and Slutzky's lemma, due to

$$\Sigma(x) = \sigma^2\, \hat{\mathbf{H}}^{-2}(x)\, \mathrm{diag}(V_N(x), W_N), \quad \forall x \in (0,1).$$

On the other hand, the second is $o_P(1)$ by Lemmas 2.7, 2.8 and 2.11. This concludes the proof of Theorem 2.2. □

2.4.4 Uniform confidence bands: proof of Theorem 2.3

We require consistency and rate of convergence of the normalized estimators of the Hessian matrix.

Lemma 2.12. *Under the Assumptions 2.1 – 2.4 we have that*

$$\begin{aligned} &\|\hat V_H - V_H\|_\infty = O_P(h) + O_P(\sqrt{\log(n)}/nh), \quad \|\hat V_N - V_N\|_\infty = O_P(\sqrt{\log(n)}/nh),\\ &\|\hat W_H - W_H\|_\infty = O_P(h) + O_P(\sqrt{\log(n)}/nh), \quad \|\hat{\mathbf{H}}_n^{-1} - \mathbf{H}^{-1}\|_\infty = O_P(h) + O_P(\sqrt{\log(n)}/nh). \end{aligned}$$

The proof is provided in Section 2.6.5.

Next, we extend our notation by incorporating the following definition. The *Lévy-concentration function* of a random variable X is given by

$$\mathbb{L}(X,\zeta) = \sup_{x\in\mathbb{R}} P(|X-x| \le \zeta), \quad \zeta \ge 0.$$

We introduce the normalized score process and its Gaussian approximation

$$\begin{aligned} Z_n^{\phi}(x) &= \sigma^{-1} V_N^{-1/2}(x)\, n\, h \left(\partial_w \mathbb{M}_n(0,\psi(x);x) - \partial_w \hat{\mathbb{M}}_n(0,\psi(x);x)\right),\\ Z_{n,G}^{\phi}(x) &= V_N^{-1/2}(x) \int_{\mathbb{R}^2} \left\langle (\nabla K)\left(\mathbf{D}_{-\psi(x)}(\mathbf{p}(x)/h - \mathbf{z})\right), (\sin\psi(x), \cos\psi(x))^T \right\rangle \mathrm{d}W(\mathbf{z}), \end{aligned} \tag{2.22}$$

where W is a Wiener sheet on $\mathbb{R}^2$. It easily follows with Lemma 2.17

$$\partial_w \hat{\mathbb{M}}_n(0,\psi(x);x) = (n^2h^2)^{-1} \sum_{i_1,i_2=1}^{n} Y_{i_1,i_2} \left\langle (\nabla K)\left(h^{-1}\mathbf{D}_{-\psi(x)}(\mathbf{p}(x) - \mathbf{x}_{\mathbf{i_1},\mathbf{i_2}})\right), (\sin\psi(x), \cos\psi(x))^T \right\rangle,$$

such that

$$Z_n^{\phi}(x) = \frac{1}{nh\sqrt{V_N(x)}\sigma} \sum_{i_1,i_2=1}^{n} \boldsymbol{\varepsilon}_{i_1,i_2} \left\langle (\nabla K)\left(h^{-1}\mathbf{D}_{-\psi(x)}(\mathbf{p}(x) - \mathbf{x}_{\mathbf{i_1},\mathbf{i_2}})\right), (\sin\psi(x), \cos\psi(x))^T \right\rangle.$$

The score process $\tilde Z_n^{\phi}$ in (2.9) can be obtained from the latter display by replacing the actual parameters (ϕ,ψ) by their estimates $(\hat\phi_n,\hat\psi_n)$ and the noise of the observations $\boldsymbol{\varepsilon}_{i_1,i_2}$ by an independent sequence of noise ξ_{i_1,i_2}. To see this, note that $\hat V_N$ in (2.7) is defined by plugging in the estimates $(\hat\phi_n,\hat\psi_n)$ for (ϕ,ψ) in the definition of V_N in (2.4).

In the spirit of (2.11) set

$$\mathbf{M}_{\phi} = \| Z_{n,G}^{\phi} \|_\infty. \tag{2.23}$$

Analogously, for the jump-slope we introduce

$$\begin{aligned} Z_n^{\psi}(x) &= \sigma^{-1} W_N^{-1/2}\, n\, h \left(\partial_\psi \mathbb{M}_n(0,\psi(x);x) - \partial_\psi \hat{\mathbb{M}}_n(0,\psi(x);x)\right),\\ Z_{n,G}^{\psi}(x) &= h^{-1} W_N^{-1/2} \int_{\mathbb{R}^2} \left\langle (\nabla K)\left(\mathbf{D}_{-\psi(x)}(\mathbf{p}(x)/h - \mathbf{z})\right), \mathbf{D}_{3\pi/2-\psi(x)}((x,\mathbf{p}(x))^T - \mathbf{z}) \right\rangle \mathrm{d}W(\mathbf{z}), \end{aligned} \tag{2.24}$$

and finally

$$\mathbf{M}_{\psi} = \| Z_{n,G}^{\psi} \|_\infty. \tag{2.25}$$

Note by Lemma 2.17 that Z_n^{ψ} and $\tilde Z_n^{\psi}$ in (2.10) are related in a similar way as Z_n^{ϕ} and $\tilde Z_n^{\phi}$.

Lévy-anti-concentration and bootstrap quantile

Lemma 2.13. *Under the Assumptions 2.1 – 2.4, on a suitable probability space there exists a Wiener sheet W such that for sufficiently large n*

$$|| Z_n^{\phi} - Z_{n,G}^{\phi} ||_\infty = O_P\Big(\frac{(\log n)^{1/2}}{n^{1/2}h}\Big).$$

Moreover, $\mathbf{M}_\phi = O_P\big((\log n)^{1/2}\big)$ *and for any sequence* $\delta_n = o((\log n)^{-1/2})$ *we have that*

$$\mathbb{L}(\mathbf{M}_\phi, \delta_n) = o(1).$$

The proof of this lemma is provided in Section 2.6.4.

Lemma 2.14. *Under the Assumptions 2.1 – 2.4, for any* $\alpha \in (0,1)$

$$\lim_n P\Big(\mathbf{M}_\phi \le q_{1-\alpha}(\tilde{\mathbf{M}}_\phi)\Big) = 1 - \alpha,$$

where $\tilde{\mathbf{M}}_\phi$ *is given in (2.11). Moreover,* $q_{1-\alpha}(\tilde{\mathbf{M}}_\phi) \cong (\log n)^{1/2}$.

The proof is given in Section 2.6.6. For the jump-slope we have analogous results.

Lemma 2.15. *Under the Assumptions 2.1 – 2.4, on a suitable probability space there exists a Wiener sheet W such that for sufficiently large n*

$$|| Z_n^{\psi} - Z_{n,G}^{\psi} ||_\infty = O_P\Big(\frac{(\log n)^{1/2}}{n^{1/2}h}\Big).$$

Moreover, $\mathbf{M}_\psi = O_P\big((\log n)^{1/2}\big)$ *and for any sequence* $\delta_n = o((\log n)^{-1/2})$ *we have that*

$$\mathbb{L}(\mathbf{M}_\psi, \delta_n) = o(1).$$

Lemma 2.16. *Under the Assumptions 2.1 – 2.4, for any* $\alpha \in (0,1)$

$$\lim_n P\Big(\mathbf{M}_\psi \le q_{1-\alpha}(\tilde{\mathbf{M}}_\psi)\Big) = 1 - \alpha,$$

where $\tilde{\mathbf{M}}_\psi$ *is given in (2.11). Moreover,* $q_{1-\alpha}(\tilde{\mathbf{M}}_\psi) \cong (\log n)^{1/2}$.

The proof of Lemma 2.15 is provided in Section 2.6.4, while the proof for Lemma 2.16 is given in Section 2.6.6.

Proof of Theorem 2.3. In the following we prove the statement for the jump-location-curve, while for the slope one can proceed analogously.

Set $\hat{d}_n = \hat{\sigma}_n q_{1-\alpha}(\tilde{\mathbf{M}}_\phi)$. Recall that for $(z_1, z_2)^T \in \mathbb{R}^2$ we write $(z)_i = z_i$ for $i = 1,2$. From (2.20) we obtain that

$$\begin{aligned}
&P\Big(\sup_{x\in I}\Big|\frac{\hat{V}_H(x)(\hat{\phi}_n(x)-\phi(x))}{(\hat{V}_N(x))^{1/2}}\Big| \ge \hat{d}_n(1+t_n)/n\Big)\\
=&P\Big(\sup_{x\in I}\Big|\frac{\hat{V}_H(x)}{(\hat{V}_N(x))^{1/2}}\, n\,h\,\Big(\hat{\mathbf{H}}_n^{-1}(x)\left(\nabla\hat{\mathbb{M}}_n(0,\psi(x);x) - \nabla\mathbb{M}_n(0,\psi(x);x) + \nabla\mathbb{M}_n(0,\psi(x);x)\right)\Big)_1\Big| \ge \hat{d}_n(1+t_n)\Big)\\
\le& P\Big(\sup_{x\in I}\Big|\frac{\hat{V}_H(x)}{(\hat{V}_N(x))^{1/2}}\, n\,h\,\Big(\hat{\mathbf{H}}_n^{-1}(x)\left(\nabla\hat{\mathbb{M}}_n(0,\psi(x);x) - \nabla\mathbb{M}_n(0,\psi(x);x)\right)\Big)_1\Big| \ge \hat{d}_n\Big)
\end{aligned}$$

$$+P\Big(\sup_{x\in I}\Big|\frac{\hat V_H(x)}{(\hat V_N(x))^{1/2}}\Big(\hat{\mathbf{H}}_n^{-1}(x)\,nh\,\nabla\mathbb{M}_n(0,\psi(x);x)\Big)_1\Big|\geq t_n\hat d_n\Big).$$

From Lemma 2.14 and the assumption on t_n we deduce that $t_n\hat d_n\to\infty$ as $n\to\infty$. Hence, by using Lemmas 2.7 and 2.12 the second term in the preceding display can be bounded by

$$P\Big(\sup_{x\in I}\Big|\frac{\hat V_H(x)}{(\hat V_N(x))^{1/2}}\Big(\hat{\mathbf{H}}_n^{-1}(x)\,nh\,\nabla\mathbb{M}_n(0,\psi(x);x)\Big)_1\Big|\geq t_n\hat d_n\Big)=o(1).$$

Recalling the notation in (2.22) the first term can be estimated by

$$\begin{aligned}
&P\Big(\sup_{x\in I}\Big|\frac{\hat V_H(x)}{(\hat V_N(x))^{1/2}}\,n\,h\,\Big(\hat{\mathbf{H}}_n^{-1}(x)\,(\nabla\hat{\mathbb{M}}_n(0,\psi(x);x)-\nabla\mathbb{M}_n(0,\psi(x);x))\Big)_1\Big|\geq\hat d_n\Big)\\
&\leq P\Big(\sup_{x\in I}\Big|\Big(\big(\frac{\hat V_H(x)}{(\hat V_N(x))^{1/2}}\hat{\mathbf{H}}_n^{-1}(x)-\frac{V_H(x)}{(V_N(x))^{1/2}}\mathbf{H}^{-1}(x)\big)\,n\,h\,(\nabla\hat{\mathbb{M}}_n(0,\psi(x);x)-\nabla\mathbb{M}_n(0,\psi(x);x))\Big)_1\Big|\geq\log(n)^{-1}\Big)\\
&\quad+P\Big(\sup_{x\in I}|Z_n^{\phi}(x)|\geq(\hat d_n-\log(n)^{-1})/\sigma\Big).
\end{aligned}\tag{2.26}$$

Lemma 2.12 implies that

$$\|\frac{\hat V_H}{\sqrt{\hat V_N}}\hat{\mathbf{H}}_n^{-1}-\frac{V_H}{\sqrt{V_N}}\mathbf{H}^{-1}\|_\infty=O_P(h)+O_P\big((\log n)^{1/2}/nh\big),$$

and together with Proposition 2.9 the first term in (2.26) can be bounded by

$$O_P\big((\log n)^{3/2}\,h\big)+O_P\big((\log n)^2/n\,h\big)=o_P(1),$$

due to Assumption 2.4. As for the second term, plugging in $\hat d_n=\hat\sigma_n q_{1-\alpha}(\tilde{\mathbf{M}}_\phi)$ gives

$$\begin{aligned}
&P\Big(\|Z_n^{\phi}\|_\infty\geq(q_{1-\alpha}(\tilde{\mathbf{M}}_\phi)\,\hat\sigma_n-\log(n)^{-1})/\sigma\Big)\\
&\leq P\Big(\|Z_n^{\phi}\|_\infty-\mathbf{M}_\phi\geq\log(n)^{-1}/\sigma\Big)+P\Big(\mathbf{M}_\phi\geq(q_{1-\alpha}(\tilde{\mathbf{M}}_\phi)\,\hat\sigma_n-2\log(n)^{-1})/\sigma\Big)\\
&\leq o(1)+P\Big(\mathbf{M}_\phi\geq q_{1-\alpha}(\tilde{\mathbf{M}}_\phi)\,\hat\sigma_n/\sigma\Big)+\mathbb{L}\big(\mathbf{M}_\phi,2\log(n)^{-1}/\sigma\big),
\end{aligned}$$

by using for the last line the first part of Lemma 2.13 together with the choice of h, and the definition of the Lévy-concentration function. The last term in this display is $o(1)$ by using the second part of Lemma 2.13. Finally, observe that for s_n as in the assumption of the theorem it holds

$$\begin{aligned}
P\Big(\mathbf{M}_\phi\geq q_{1-\alpha}(\tilde{\mathbf{M}}_\phi)\,\hat\sigma_n/\sigma\Big)&\leq P\Big(\mathbf{M}_\phi\geq q_{1-\alpha}(\tilde{\mathbf{M}}_\phi)(1+s_n)\Big)+P\big(\big|\hat\sigma_n/\sigma-1\big|\geq s_n\big)\\
&\leq P\Big(\mathbf{M}_\phi\geq q_{1-\alpha}(\tilde{\mathbf{M}}_\phi)(1+s_n)\Big)+o(1)\\
&\leq P\big(\mathbf{M}_\phi\geq q_{1-\alpha}(\tilde{\mathbf{M}}_\phi)\big)+\mathbb{L}\big(\mathbf{M}_\phi,s_n\,q_{1-\alpha}(\tilde{\mathbf{M}}_\phi)\big)+o(1)\\
&=P\Big(\mathbf{M}_\phi\geq q_{1-\alpha}(\tilde{\mathbf{M}}_\phi)\Big)+o(1),
\end{aligned}$$

where we used for the last line the second statement of Lemma 2.13 together with the fact that by Lemma 2.14, $s_nq_{1-\alpha}(\tilde{\mathbf{M}}_\phi)=o((\log n)^{-1/2})$. Summarizing, we obtain that

$$P\Big(\sup_{x\in I}\Big|\frac{\hat V_H(x)(\hat\phi_n(x)-\phi(x))}{(\hat V_N(x))^{1/2}}\Big|\geq\hat d_n(1+t_n)/n\Big)\leq P\big(\mathbf{M}_\phi\geq q_{1-\alpha}(\tilde{\mathbf{M}}_\phi)\big)+o(1)=\alpha+o(1),$$

where in the final step we used Lemma 2.14. □

2.5 Discussion

In this chapter we developed methods to construct asymptotic confidence sets for the jump curve in an otherwise smooth two-dimensional regression function, for which to the best of our knowledge no methods were previously available. Additionally, this work offers various extensions and issues which we will discuss here in the following.

Optimality

From Theorem 4.3 below, it follows that the estimate $\hat{\phi}_n$ for the jump-location-curve attains the optimal rate of convergence up to a logarithmic factor. In addition, Theorem 4.3 shows also that higher smoothness assumptions on the jump-location-curve ϕ or the continuous background picture m do not affect the minimax optimal rate for estimation of ϕ. Hence, the approach considered here is optimal (up to a logarithmic factor) for those higher smoothness-cases as well. Furthermore, adaptivity considerations as in Section 1.4 for these confidence bands are not sensible, as the order of the width of the confidence bands is not affected by higher smoothness of the jump-location-curve.

Multivariate setting

An extension to the multivariate setting, especially to three dimensions would certainly be relevant. The definition of the estimates for this case is similar to Müller and Song (1994) leading to $d-1$ estimates for the slopes, if the model dimension is $d \in \mathbb{N}$. Presumably the crucial points would be the assumption on the kernels and the deviation of the asymptotic limit function as in (2.18). Using techniques as in Proksch et al. (2015) one should obtain a multivariate Gaussian approximation similar as in Lemma A.1, while multivariate extensions of the other auxiliary results are straightforward. Thus, it is reasonable to believe that the results of this chapter can be extended to the multivariate case and the rates of convergence adapt to the multivariate setting in the "usual" way, i.e. the rate of convergence for the location of the jump curve is of order $O({}^{\log(n_0)^{1/2}}/{}_{(n_0)^{1/d}})$, where $n_0 = n^d$ is the total number of design points, and the rates of convergence for the $d-1$ different slopes of the jump curve are $O({}^{\log(n_0)^{1/2}}/{}_{(n_0 h)^{1/d}})$.

Several jump curves

It is also of interest to consider for the regression function an extension to several jump-location-curves in (2.2), that is for some $J \in \mathbb{N}$

$$m_\phi(x,y) = m(x,y) + \sum_{k=1}^{J} j_{k,\tau,\phi}(x,y),$$
$$j_{k,\tau,\phi}(x,y) = \tau_k(x) 1_{[0,\phi_k(x)]}(y), \quad k = 1,\dots,J$$

where $m : [0,1]^2 \to \mathbb{R}$ is the smooth part of the image, $\tau_k : [0,1] \to \mathbb{R}_+$ are the jump-height curves and $\phi_k : [0,1] \to (0,1)$ the jump-location-curves. In order to guarantee identification one has to assume that the images of the jump curves are well separated, i.e. $B_\rho(\phi_i[0,1]) \cap B_\rho(\phi_j[0,1]) = \emptyset$ for any $i \neq j, i, j \in \{1,\dots,J\}$, where $B_\rho(A)$ denotes the ρ-neighborhood of a set $A \subset \mathbb{R}$ and $\rho > 0$ is some appropriate constant. In particular, for finite sample analysis the images have to be separated uniformly by a suitable multiple of the bandwidth. Theoretical concepts should be transferable with some additional notational effort and confidence bands could be constructed based on a Bonferroni-correction.

Alternative approach

Following the idea of Mammen and Polonik (2013) which is defining a confidence band by a level-set-condition based on the empirical contrast will lead to an alternative approach for constructing uniform confidence bands for the jump-location-curve ϕ in model (2.1). This approach is sensible as the maximal empirical contrast builds a ridge around the jump-curve, where the height of the ridge is $\hat{\tau}_n$ (see (1.16)). Thus, choosing the confidence bands as a uniform neighborhood of the ridge position should lead to reasonable confidence bands. Indeed, this approach works by using the quantile of $\tilde{M}_\tau := \sup_{(x,w,\psi)^T\in\Theta_n} |\tilde{Z}_n^\tau(x,w,\psi)|$ to define such a neighborhood (see Mammen and Polonik (2013) for a rough formulation), where

$$\tilde{Z}_n^\tau(x,w,\psi) := \frac{1}{nh}\sum_{i_1,i_2=1}^{n} \xi_{i_1,i_2} K\left(h^{-1}\mathbf{D}_{-\psi}\left(\mathbf{p}(x)+hw\mathbf{e_2}-\mathbf{x_{i_1,i_2}}\right)\right),$$

and $\xi_{1,1},\ldots,\xi_{n,n}$ are independent and standard normally distributed random variables which are independent of Y_{i_1,i_2} as well. The quantiles may be determined by simulations. Unfortunately, this approach leads to confidence bands which are asymptotically wider than those in (2.12). To be more precise, the width of these asymptotic confidence bands will be of the same order as in (2.12) inflated by the factor h^{-1} coming from the estimation rate for τ, see Theorem 2.1.

Indirect observations

Apart from noise, images are often observed with blurring, that is, after convolution with a point-spread function Υ. The blurred version of model (2.1) is

$$Y_{i_1,i_2} = (\Upsilon * m_\phi)(\mathbf{x_{i_1,i_2}}) + \varepsilon_{i_1,i_2}, \quad (i_1,i_2)\in\{1,\ldots,n\}^2, \tag{2.27}$$

where $(\Upsilon * m_\phi)(\mathbf{x}) = \int\int \Upsilon(\mathbf{x}-\mathbf{y})m_\phi(\mathbf{y})\,d\mathbf{y}$ for a point-spread function $\Upsilon:\mathbb{R}^2\to\mathbb{R}$. Considering the literature, there are very few papers dealing with this type of problem. Goldenshluger and Spokoiny (2006) estimated the unique discontinuity curve along a convex set of an otherwise smooth image under observation of the Radon transform of the image, while Kang and Qiu (2014) proposed a jump-curve-detection-method for a model similar to (2.27) based on local linear kernel smoothing. Thus, there seems to be a lot of potential for future research on statistical inference for this indirect problem.

2.6 Detailed proofs

2.6.1 Order of the asymptotic bias: proof of Lemma 2.7

Recall from (2.5) that the rescaled contrast function is given by

$$\hat{\mathbb{M}}_n(w,\psi;x) = (nh)^{-2}\sum_{i_1,i_2=1}^{n} Y_{i_1,i_2} K\left(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}(x)+wh\mathbf{e_2}-\mathbf{x_{i_1,i_2}})\right), \qquad \mathbf{p}(x)=(x,\phi(x))^T. \tag{2.28}$$

Let

$$T(\mathbf{z};w,\psi) = \mathbf{D}_{3\pi/2-\psi}(\mathbf{p}(x)+wh\mathbf{e_2}-\mathbf{z}),$$

so that

$$T(\mathbf{p}(x)+wh\mathbf{e_2}-h\mathbf{D}_\psi\mathbf{z};w,\psi)=h\,\mathbf{D}_{3\pi/2}\mathbf{z}=h\,(z_2,-z_1)^T. \tag{2.29}$$

Lemma 2.17. *It holds that*

$$\partial_w\hat{\mathbb{M}}_n(w,\psi;x)=(n^2h^2)^{-1}\sum_{i_1,i_2=1}^{n}Y_{i_1,i_2}\left\langle(\nabla K)\left(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}(x)+wh\mathbf{e_2}-\mathbf{x_{i_1,i_2}})\right),(\sin(\psi),\cos(\psi))^T\right\rangle,$$
$$\partial_\psi\hat{\mathbb{M}}_n(w,\psi;x)=(n^2h^3)^{-1}\sum_{i_1,i_2=1}^{n}Y_{i_1,i_2}\left\langle(\nabla K)\left(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}(x)+wh\mathbf{e_2}-\mathbf{x_{i_1,i_2}})\right),T(\mathbf{x_{i_1,i_2}};w,\psi))\right\rangle.$$

Proof of Lemma 2.17. We have that

$$\partial_w h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}(x)+wh\mathbf{e_2}-\mathbf{x_{i_1,i_2}})=(\sin(\psi),\cos(\psi))^T,$$
$$\partial_\psi h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}(x)+wh\mathbf{e_2}-\mathbf{x_{i_1,i_2}})=h^{-1}\mathbf{D}_{3/2\pi-\psi}(\mathbf{p}(x)+wh\mathbf{e_2}-\mathbf{x_{i_1,i_2}})=h^{-1}T(\mathbf{x_{i_1,i_2}};w,\psi).$$

The statement follows together with (2.28) by using the chain rule. □

Lemma 2.18. *Under the Assumption 2.3 it holds that*

(i) $\int_{[-1,1]^2}K^{(0,1)}(z_1,z_2)\,z_1\,\mathrm{d}z_1\mathrm{d}z_2=\int_{[-1,1]^2}K^{(1,0)}(z_1,z_2)\,z_2\,\mathrm{d}z_1\mathrm{d}z_2=0;$

(ii) $\int_{[-1,1]\times[0,1]}K^{(1,0)}(z_1,z_2)\;z_1\,\mathrm{d}z_1\mathrm{d}z_2=\int_{[-1,1]\times[0,1]}K^{(0,1)}(z_1,z_2)\;z_2\,\mathrm{d}z_1\mathrm{d}z_2=-1;$

(iii) $\int_{[-1,1]\times[0,1]}K^{(0,1)}(z_1,z_2)\;z_1\,\mathrm{d}z_1\mathrm{d}z_2=\int_{[-1,1]\times[0,1]}K^{(1,0)}(z_1,z_2)\;z_2\,\mathrm{d}z_1\mathrm{d}z_2=0;$

(iv) $\int_{[-1,1]^2}K^{(1,0)}(z_1,z_2)\;z_1\;z_2\,\mathrm{d}z_1\mathrm{d}z_2=\int_{[-1,1]^2}K^{(0,1)}(z_1,z_2)\;z_1\;z_2\,\mathrm{d}z_1\mathrm{d}z_2=0;$

(v) $\int_{[-1,1]^2}K^{(1,0)}(z_1,z_2)\,z_2^2\,\mathrm{d}z_1\mathrm{d}z_2=\int_{[-1,1]^2}K^{(0,1)}(z_1,z_2)\,z_1^2\,\mathrm{d}z_1\mathrm{d}z_2=0;$

(vi) $\int_{[-1,1]\times[0,1]}K^{(1,0)}(z_1,z_2)\;z_1\;z_2\,\mathrm{d}z_1\mathrm{d}z_2=\int_{[-1,1]\times[0,1]}K^{(0,1)}(z_1,z_2)\;z_1\;z_2\,\mathrm{d}z_1\mathrm{d}z_2=0;$

(vii) $\int_{[-1,1]\times[0,1]}K^{(1,0)}(z_1,z_2)\;z_1^2\,\mathrm{d}z_1\mathrm{d}z_2=\int_{[-1,1]\times[0,1]}K^{(0,1)}(z_1,z_2)\;z_2^2\,\mathrm{d}z_1\mathrm{d}z_2=0.$

Proof of Lemma 2.18. Some equalities follow by symmetry and normalization of K_1 and K_2, others require the boundary properties. For example, by integration by parts

$$\int_{-1}^{1}K_1^{(1)}(x)\;x\,\mathrm{d}x=-\int_{-1}^{1}K(x)\,\mathrm{d}x=-1$$

since $K_1(1)=K_1(-1)=0$ and $\int K=1$ by assumption. In addition, $\int_0^1 K_2(x)=1$ which yield the first statement in (ii). □

Proof of Lemma 2.7. For sake of brevity we write $\mathbf{v}_{\psi(x)}=(\sin\psi(x),\cos\psi(x))^T$. We discuss the expected values of the partial derivatives of $\hat{\mathbb{M}}_n\,(0,\psi(x);x)$ respectively in the following.

Partial derivative with respect to w

Using Lemma 2.17 yields

$$n\,h\,\mathbb{E}(\,\partial_w\hat{\mathbb{M}}_n\,(0,\psi(x);x))=(nh)^{-1}\sum_{i_1,i_2=1}^{n}m(\mathbf{x_{i_1,i_2}})\left\langle(\nabla K)\left(h^{-1}\mathbf{D}_{-\psi(x)}(\mathbf{p}(x)-\mathbf{x_{i_1,i_2}})\right),\mathbf{v}_{\psi(x)}\right\rangle$$

$$
\begin{aligned}
&+ (nh)^{-1} \sum_{i_1,i_2=1}^{n} j_\tau(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}) \left\langle (\nabla K)\left(h^{-1}\mathbf{D}_{-\psi(x)}(\mathbf{p}(x)-\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}})\right), \mathbf{v}_{\psi(x)} \right\rangle \\
=:\ & S_n + J_n.
\end{aligned}
$$

Note that we omitted the dependency on x for S_n resp. J_n. From Lemma B.1 one obtains

$$
\begin{aligned}
\frac{S_n}{nh} = & \int_{[-1,1]^2} \left(m(\mathbf{p}(x)-h\mathbf{D}_{\psi(x)}\mathbf{z}) - m(\mathbf{p}(x))\right) \left\langle \nabla K(\mathbf{z}), \mathbf{v}_{\psi(x)} \right\rangle \mathrm{d}\mathbf{z} \\
& + \int_{[-1,1]^2} m(\mathbf{p}(x)) \left\langle \nabla K(\mathbf{z}), \mathbf{v}_{\psi(x)} \right\rangle \mathrm{d}\mathbf{z} + O_{x\in I}((nh)^{-1})
\end{aligned}
$$

by taking $g_1(\mathbf{z}) = m(\mathbf{z})$, $f(\mathbf{z}) = (\nabla K(\mathbf{z})\ \nabla K(\mathbf{z}))$, $g_2 \equiv \mathbf{e}_1$, $g_3 \equiv \mathbf{v}_{\psi(x)}$, such that $j = 1$, $r_1 = 2$, $r_2 = r_3 = 0$. As the components $\nabla K(\mathbf{z})$ are odd, the second term is zero. Concerning the first term, obtain by Taylor expansion

$$
\begin{aligned}
& \int_{[-1,1]^2} \left(m(\mathbf{p}(x)-h\mathbf{D}_{\psi(x)}\mathbf{z}) - m(\mathbf{p}(x))\right) \left\langle \nabla K(\mathbf{z}), \mathbf{v}_{\psi(x)} \right\rangle \mathrm{d}\mathbf{z} \\
&= h m^{(1,0)}(\mathbf{p}(x)) \int_{[-1,1]^2} \left(\mathbf{D}_{\psi(x)}\mathbf{z}\right)_1 \left\langle \nabla K(\mathbf{z}), \mathbf{v}_{\psi(x)} \right\rangle \mathrm{d}\mathbf{z} + h m^{(0,1)}(\mathbf{p}(x)) \int_{[-1,1]^2} \left(\mathbf{D}_{\psi(x)}\mathbf{z}\right)_2 \left\langle \nabla K(\mathbf{z}), \mathbf{v}_{\psi(x)} \right\rangle \mathrm{d}\mathbf{z} \\
&\quad + O_{x\in I}(h^2).
\end{aligned}
$$

Both integrals on the right-hand side are zero due to Lemma 2.21, (vii) and Lemma 2.18, (i). Thus, $S_n = O_{x\in I}(nh^3) + O_{x\in I}(1)$. Similarly, with the same functions f, g_2, g_3 as above for Lemma B.2 obtain

$$
\begin{aligned}
\frac{J_n}{nh} = & \int_{h^{-1}\mathbf{D}_{-\psi(x)}\left(\mathbf{p}(x)-[0,1]^2\backslash\mathrm{epi}(\phi)\right)} \left(\tau\left(x-(h\mathbf{D}_{\psi(x)}\mathbf{z})_1\right) - \tau(x)\right) \left\langle \nabla K(\mathbf{z}), \mathbf{v}_{\psi(x)} \right\rangle \mathrm{d}\mathbf{z} \\
& + \int_{h^{-1}\mathbf{D}_{-\psi(x)}\left(\mathbf{p}(x)-[0,1]^2\backslash\mathrm{epi}(\phi)\right)} \tau(x) \left\langle \nabla K(\mathbf{z}), \mathbf{v}_{\psi(x)} \right\rangle \mathrm{d}\mathbf{z} + O_{x\in I}((nh)^{-1}).
\end{aligned}
$$

Without loss of generality let h be so small (n large enough) such that

$$
[-1,1]\times[0,1] \subset h^{-1}\mathbf{D}_{-\psi(x)}\left(\mathbf{p}(x) - [0,1]^2\backslash\mathrm{epi}(\phi)\right). \tag{2.30}
$$

Note that $\int_{[-1,1]\times[0,1]} \left\langle \nabla K(\mathbf{z}), \mathbf{v}_{\psi(x)} \right\rangle \mathrm{d}\mathbf{z} = 0$, (e.g. Lemma 2.21 (xi)) such that the second term is zero in the latter display. For the first term one has by Taylor expansion

$$
\begin{aligned}
& \int_{[-1,1]\times[0,1]} \left(\tau\left(x-(h\mathbf{D}_{\psi(x)}\mathbf{z})_1\right) - \tau(x)\right) \left\langle \nabla K(\mathbf{z}), \mathbf{v}_{\psi(x)} \right\rangle \mathrm{d}\mathbf{z} \\
& \qquad = h\tau^{(1)}(x) \int_{[-1,1]\times[0,1]} (\mathbf{D}_{\psi(x)}\mathbf{z})_1 \left\langle \nabla K(\mathbf{z}), \mathbf{v}_{\psi(x)} \right\rangle \mathrm{d}\mathbf{z} + O_{x\in I}(h^2),
\end{aligned}
$$

where the integral can be further simplified by Lemma 2.18, (ii) and (iii) to

$$
\begin{aligned}
& \int_{[-1,1]\times[0,1]} (\mathbf{D}_{\psi(x)}\mathbf{z})_1 \left\langle \nabla K(\mathbf{z}), \mathbf{v}_{\psi(x)} \right\rangle \mathrm{d}\mathbf{z} \\
&= \int_{[-1,1]\times[0,1]} \left(\cos(\psi(x))\, z_1 + \sin(-\psi(x))\, z_2\right) \left(\sin(\psi(x))\, K^{(1,0)}(z_1,z_2) + \cos(\psi(x))\, K^{(0,1)}(z_1,z_2)\right) \mathrm{d}z_1 \mathrm{d}z_2 \\
&= -\cos(\psi(x))\sin(\psi(x)) - \cos(\psi(x))\sin(-\psi(x)) = 0.
\end{aligned}
$$

In summary, $J_n = O_{x\in I}(nh^3) + O_{x\in I}(1)$.

Partial derivative with respect to ψ

Using Lemma 2.17 yields

$$\begin{aligned}
n\,h\,\mathbb{E}(\,\partial_\psi \hat{\mathbb{M}}_n(0,\psi(x);x)) &= (nh^2)^{-1}\sum_{i_1,i_2=1}^{n} m(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}})\left\langle (\nabla K)(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}(x)-\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}})), T(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}};0,\psi(x)))\right\rangle \\
&\quad + (nh^2)^{-1}\sum_{i_1,i_2=1}^{n} j_\tau(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}})\left\langle (\nabla K)(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}(x)-\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}})), T(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}};0,\psi(x)))\right\rangle \\
&=: S_n + J_n.
\end{aligned}$$

Lemma B.1 gives us together with (2.29) that uniformly over x

$$\begin{aligned}
\frac{S_n}{nh} &= \int_{[-1,1]^2}\left(m(\mathbf{p}(x)-h\mathbf{D}_{\psi(x)}(z_1,z_2)^T)-m(\mathbf{p}(x))\right)\left\langle \nabla K(z_1,z_2),(z_2,-z_1)^T\right\rangle \mathrm{d}z_1\mathrm{d}z_2 \\
&\quad + \int_{[-1,1]^2} m(\mathbf{p}(x))\left\langle \nabla K(z_1,z_2),(z_2,-z_1)^T\right\rangle \mathrm{d}z_1\mathrm{d}z_2 + O_{x\in I}((nh)^{-1}),
\end{aligned}$$

by taking g_1, g_2 and f as before and $g_3(\mathbf{z}) = T(\mathbf{z};0,\psi(x)))$, so that $j=1$, $r_1=3$, $r_2=0$ and $r_3=1$. The second term on the right-hand side of the latter display is zero since the functions $x\mapsto K_1^{(1)}(x)$ and $x\mapsto K_1(x)\,x$ are odd. By Taylor expansion in the first term

$$\begin{aligned}
&\int_{[-1,1]^2}\left(m(\mathbf{p}(x)-h\mathbf{D}_{\psi(x)}(z_1,z_2)^T)-m(\mathbf{p}(x))\right)\left\langle \nabla K(z_1,z_2),(z_2,-z_1)^T\right\rangle \mathrm{d}z_1\mathrm{d}z_2 \\
&= h\,m^{(1,0)}(\mathbf{p}(x))\int_{[-1,1]^2}\left(\mathbf{D}_{\psi(x)}(z_1,z_2)^T\right)_1\left\langle \nabla K(z_1,z_2),(z_2,-z_1)^T\right\rangle \mathrm{d}z_1\mathrm{d}z_2 \\
&\quad + h\,m^{(0,1)}(\mathbf{p}(x))\int_{[-1,1]^2}\left(\mathbf{D}_{\psi(x)}(z_1,z_2)^T\right)_2\left\langle \nabla K(z_1,z_2),(z_2,-z_1)^T\right\rangle \mathrm{d}z_1\mathrm{d}z_2 + O_{x\in I}(h^2).
\end{aligned}$$

Lemma 2.18, (iv) and (v) imply that both integrals on the right-hand side are zero. Therefore, one has $S_n = O_{x\in I}(nh^3) + O_{x\in I}(1)$. We still assume that h is so small respectively n is so large that (2.30) holds. Thus, application of Lemma B.2 with the same functions g_2, g_2 and f as just yields

$$\begin{aligned}
\frac{J_n}{nh} &= \int_{[-1,1]\times[0,1]}\left(\tau\left(x-(h\mathbf{D}_{\psi(x)}\mathbf{z})_1\right)-\tau(x)\right)\left\langle \nabla K(z_1,z_2),(z_2,-z_1)^T\right\rangle \mathrm{d}z_1\mathrm{d}z_2 \\
&\quad + \int_{[-1,1]\times[0,1]} \tau(x)\left\langle \nabla K(z_1,z_2),(z_2,-z_1)^T\right\rangle \mathrm{d}z_1\mathrm{d}z_2 + O_{x\in I}((nh)^{-1}).
\end{aligned}$$

Since $\int_{[-1,1]\times[0,1]}\langle \nabla K(z_1,z_2),(z_2,-z_1)\rangle\,\mathrm{d}z_1\mathrm{d}z_2 = 0$, by Lemma 2.18, (iii), we only discuss the first term on the right-hand side of the latter display. Similarly as before by a Taylor expansion,

$$\begin{aligned}
&\int_{[-1,1]\times[0,1]}\left(\tau\left(x-(h\mathbf{D}_{\psi(x)}\mathbf{z})_1\right)-\tau(x)\right)\left\langle \nabla K(z_1,z_2),(z_2,-z_1)^T\right\rangle \mathrm{d}z_1\mathrm{d}z_2 \\
&= h\,\tau^{(1)}(x)\int_{[-1,1]\times[0,1]}(\mathbf{D}_{\psi(x)}\mathbf{z})_1\left\langle \nabla K(z_1,z_2),(z_2,-z_1)^T\right\rangle \mathrm{d}z_1\mathrm{d}z_2 + O_{x\in I}(h^2),
\end{aligned}$$

where by Lemma 2.18, (vi) and (vii), the integral is equivalent to

$$\begin{aligned}
&\int_{[-1,1]\times[0,1]}(\mathbf{D}_{\psi(x)}\mathbf{z})_1\left\langle \nabla K(z_1,z_2),(z_2,-z_1)^T\right\rangle \mathrm{d}z_1\mathrm{d}z_2 \\
&= \int_{[-1,1]\times[0,1]}\left(\cos(\psi(x))\,z_1+\sin(-\psi(x))\,z_2\right)\left(K^{(1,0)}(z_1,z_2)\,z_2-K^{(0,1)}(z_1,z_2)\,z_1\right)\mathrm{d}z_1\mathrm{d}z_2 = 0.
\end{aligned}$$

Consequently, $J_n = O_{x\in I}(nh^3) + O_{x\in I}(1)$.

All in all, by Assumption 2.4 on h,

$$nh\,||\,\mathbb{E}\big(\nabla\hat{\mathbb{M}}_n\,(0,\psi(x);x)\big)\,|| = O_{x\in I}(nh^3) + O_{x\in I}(1) = O_{x\in I}(1)$$

as asserted. □

2.6.2 Convergence of the Hessian matrix: proofs of Lemma 2.8 and Lemma 2.11

Similarly to Lemma 2.17 it is straightforward to calculate the second derivatives.

Lemma 2.19. *It holds for any $x \in [0,1]$,*

$$\begin{aligned}
&\partial_w^2\hat{\mathbb{M}}_n(w,\psi;x)\\
&\quad= (n^2h^2)^{-1}\sum_{i_1,i_2=1}^{n} Y_{i_1,i_2}\Big\langle (\nabla\nabla^T K)\Big(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}(x)+wh\mathbf{e_2}-\mathbf{x_{i_1,i_2}})\Big)(\sin(\psi),\cos(\psi))^T,(\sin(\psi),\cos(\psi))^T\Big\rangle,\\
&\partial_\psi\partial_w\hat{\mathbb{M}}_n(w,\psi;x)\\
&\quad= (n^2h^3)^{-1}\sum_{i_1,i_2=1}^{n} Y_{i_1,i_2}\Big\langle (\nabla\nabla^T K)\Big(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}(x)+wh\mathbf{e_2}-\mathbf{x_{i_1,i_2}})\Big)(\sin(\psi),\cos(\psi))^T,T(\mathbf{x_{i_1,i_2}};w,\psi)\Big\rangle\\
&\qquad+(n^2h^2)^{-1}\sum_{i_1,i_2=1}^{n} Y_{i_1,i_2}\left\langle (\nabla K)\Big(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}(x)+wh\mathbf{e_2}-\mathbf{x_{i_1,i_2}})\Big)^T,(\cos(\psi),-\sin(\psi))\right\rangle,\\
&\partial_\psi^2\hat{\mathbb{M}}_n(w,\psi;x)\\
&\quad= (n^2h^4)^{-1}\sum_{i_1,i_2=1}^{n} Y_{i_1,i_2}\Big\langle (\nabla\nabla^T K)\Big(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}(x)+wh\mathbf{e_2}-\mathbf{x_{i_1,i_2}})\Big)T(\mathbf{x_{i_1,i_2}};w,\psi(x)),T(\mathbf{x_{i_1,i_2}};w,\psi)\Big\rangle\\
&\qquad-(n^2h^3)^{-1}\sum_{i_1,i_2=1}^{n} Y_{i_1,i_2}\left\langle (\nabla K)\Big(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}(x)+wh\mathbf{e_2}-\mathbf{x_{i_1,i_2}})\Big)^T,\mathbf{D}_{-3\pi/2}\,T(\mathbf{x_{i_1,i_2}};w,\psi)\right\rangle.
\end{aligned}$$

We need the following two auxiliary results to verify the stochastic convergence of the Hessian matrix.

Lemma 2.20. *Let $d \in \mathbb{N}$ and $(\hat{f}_n)_n : \mathbb{R}^d \times I \to \mathbb{R}$ be random functions, which are uniformly Lipschitz, i.e.*

$$\big|\hat{f}_n(\mathbf{z}_1;x) - \hat{f}_n(\mathbf{z}_2;x)\big| \le L_n\,|\mathbf{z}_1 - \mathbf{z}_2|,$$

where $L_n = O_P(1)$ uniformly over I. Furthermore, assume there exists $\eta_0 : I \to \mathbb{R}^d$ such that the functions $\hat{f}_n$ are weakly uniformly consistent with the function $f : \mathbb{R}^d \times I \to \mathbb{R}$ at η_0:

$$\sup_{x\in I}|\hat{f}_n(\eta_0(x);x) - f(\eta_0(x);x)| \xrightarrow{P} 0.$$

Then for each sequence of random functions $(\hat{\eta}_n)_{n\in\mathbb{N}}$ with $\hat{\eta}_n : I \to \mathbb{R}^d$ such that $||\,\hat{\eta}_n - \eta_0\,||_\infty \xrightarrow{P} 0$ it holds that

$$\sup_{x\in I}|\hat{f}_n(\hat{\eta}_n(x);x) - f(\eta_0(x);x)| \xrightarrow{P} 0.$$

Proof of Lemma 2.20.

$$\sup_{x\in I}\big|\hat{f}_n(\hat{\eta}_n(x);x) - f(\eta_0(x);x)\big| \le \sup_{x\in I}\big|\hat{f}_n(\hat{\eta}_n(x);x) - \hat{f}_n(\eta_0(x);x)\big| + \sup_{x\in I}\big|\hat{f}_n(\eta_0(x);x) - f(\eta_0(x);x)\big|$$

$$\leq L_n \sup_{x\in I} |\hat{\eta}_n(x)-\eta_0(x)| + o_P(1) = o_P(1).$$

□

Lemma 2.21. *Under the Assumption 2.3 it holds that*

(i) $\int_{[-1,1]^2} K^{(2,0)}(\mathbf{z})\,\mathrm{d}\mathbf{z} = \int_{[-1,1]^2} K^{(1,1)}(\mathbf{z})\,\mathrm{d}\mathbf{z} = \int_{[-1,1]^2} K^{(0,2)}(\mathbf{z})\,\mathrm{d}\mathbf{z} = 0;$

(ii) $\int_{[-1,1]^2} K^{(2,0)}(z_1,z_2)\,z_2\,\mathrm{d}z_1\mathrm{d}z_2 = \int_{[-1,1]^2} K^{(1,1)}(z_1,z_2)\,z_1\,\mathrm{d}z_1\mathrm{d}z_2 = 0;$

(iii) $\int_{[-1,1]^2} K^{(1,1)}(z_1,z_2)\,z_2\,\mathrm{d}z_1\mathrm{d}z_2 = \int_{[-1,1]^2} K^{(0,2)}(z_1,z_2)\,z_1\,\mathrm{d}z_1\mathrm{d}z_2 = 0;$

(iv) $\int_{[-1,1]^2} K^{(1,0)}(\mathbf{z})\,\mathrm{d}\mathbf{z} = \int_{[-1,1]^2} K^{(0,1)}(\mathbf{z})\,\mathrm{d}\mathbf{z} = 0;$

(v) $\int_{[-1,1]^2} K^{(2,0)}(z_1,z_2)\,z_2^2\,\mathrm{d}z_1\mathrm{d}z_2 = \int_{[-1,1]^2} K^{(1,1)}(z_1,z_2)\,z_1z_2\,\mathrm{d}z_1\mathrm{d}z_2 = 0;$

(vi) $\int_{[-1,1]^2} K^{(0,2)}(z_1,z_2)\,z_1^2\,\mathrm{d}z_1\mathrm{d}z_2 = 0;$

(vii) $\int_{[-1,1]^2} K^{(1,0)}(z_1,z_2)\,z_1\,\mathrm{d}z_1\mathrm{d}z_2 = \int_{[-1,1]^2} K^{(0,1)}(z_1,z_2)\,z_2\,\mathrm{d}z_1\mathrm{d}z_2 = 0.$

(viii) $\int_{[-1,1]\times[0,1]} K^{(2,0)}(\mathbf{z})\,\mathrm{d}\mathbf{z} = \int_{[-1,1]\times[0,1]} K^{(1,1)}(\mathbf{z})\,\mathrm{d}\mathbf{z} = 0, \quad \int_{[-1,1]\times[0,1]} K^{(0,2)}(\mathbf{z})\,\mathrm{d}\mathbf{z} = -K_2^{(1)}(0);$

(ix) $\int_{[-1,1]\times[0,1]} K^{(2,0)}(z_1,z_2)\,z_2\,\mathrm{d}z_1\mathrm{d}z_2 = \int_{[-1,1]\times[0,1]} K^{(1,1)}(z_1,z_2)\,z_1\,\mathrm{d}z_1\mathrm{d}z_2 = 0;$

(x) $\int_{[-1,1]\times[0,1]} K^{(1,1)}(z_1,z_2)\,z_2\,\mathrm{d}z_1\mathrm{d}z_2 = \int_{[-1,1]\times[0,1]} K^{(0,2)}(z_1,z_2)\,z_1\,\mathrm{d}z_1\mathrm{d}z_2 = 0;$

(xi) $\int_{[-1,1]\times[0,1]} K^{(1,0)}(\mathbf{z})\,\mathrm{d}\mathbf{z} = \int_{[-1,1]\times[0,1]} K^{(0,1)}(\mathbf{z})\,\mathrm{d}\mathbf{z} = 0;$

(xii) $\int_{[-1,1]\times[0,1]} K^{(1,1)}(z_1,z_2)\,z_1z_2\,\mathrm{d}z_1\mathrm{d}z_2 = 1, \quad \int_{[-1,1]\times[0,1]} K^{(2,0)}(z_1,z_2)\,z_2^2\,\mathrm{d}z_1\mathrm{d}z_2 = 0;$

(xiii) $\int_{[-1,1]\times[0,1]} K^{(0,2)}(z_1,z_2)\,z_1^2\,\mathrm{d}z_1\mathrm{d}z_2 = -K_2^{(1)}(0)\int_{-1}^{1} K_1(y)\,y^2\,\mathrm{d}y;$

(xiv) $\int_{[-1,1]\times[0,1]} K^{(1,0)}(z_1,z_2)\,z_1\,\mathrm{d}z_1\mathrm{d}z_2 = \int_{[-1,1]\times[0,1]} K^{(0,1)}(z_1,z_2)\,z_2\,\mathrm{d}z_1\mathrm{d}z_2 = -1.$

Proof of Lemma 2.21. Some equalities follow by symmetry and normalization of K_1 and K_2, others require the boundary properties. For example,

$$\int_0^1 K_2^{(2)}(z)\,\mathrm{d}\mathbf{z} = K_2^{(1)}(1) - K_2^{(1)}(0) = -K_2^{(1)}(0)$$

since $K_2^{(1)}(1) = 0$ by assumption, which yield the last statement in (viii). □

Proof of Lemma 2.8. We show that

$$\sup_{x\in I} ||\,\nabla\nabla^T\hat{\mathbb{M}}_n(0,\psi(x);x) - \mathbf{H}(x)\,|| = O_P(h) + O_P((nh)^{-1})$$

as well as the stochastic Lipschitz continuity of

$$\hat{f}_n(w,\psi;x) := \int_0^1 \nabla\nabla^T\hat{\mathbb{M}}_n(tw,\psi(x)+t(\psi-\psi(x));x)\,\mathrm{d}t$$

on Θ_n. Note that $\hat{f}_n(0,\psi(x);x) = \nabla\nabla^T\hat{\mathbb{M}}_n(0,\psi(x);x)$. The lemma then follows from Lemma 2.20 together with Proposition 2.4 by taking $\eta_0(x) = (0,\psi(x))$.

Uniform stochastic convergence

As we have to show stochastic convergence of a symmetric matrix, we break it down to showing stochastic convergence of the components.

For sake of brevity we write $\mathbf{v}_{\psi(x)} = (\sin\psi(x), \cos\psi(x))^T$. Also note that all the O-terms in Lemma B.1, Lemma B.2 and the o_P-term in Lemma B.3 are uniform in x, so that all occurring O-Terms in the following Steps 1 – 3 hold uniformly in x.

Step 1:

We show that

$$\sup_{x\in I} |\partial_w^2 \hat{\mathbb{M}}_n(0,\psi(x);x) + \tau(x)\cos^2(\psi(x)) K_2^{(1)}(0)| = O_P(h) + O_P((nh)^{-1}).$$

Use Lemma 2.19 and split $\partial_w^2 \hat{\mathbb{M}}_n(0,\psi(x);x)$ into three terms

$$\begin{aligned}\partial_w^2 \hat{\mathbb{M}}_n(0,\psi(x);x) &= (n^2h^2)^{-1} \sum_{i_1,i_2=1}^{n} Y_{i_1,i_2} \left\langle (\nabla\nabla^T K)\left(h^{-1}\mathbf{D}_{-\psi(x)}(\mathbf{p}(x) - \mathbf{x}_{\mathbf{i_1},\mathbf{i_2}})\right) \mathbf{v}_{\psi(x)}, \mathbf{v}_{\psi(x)} \right\rangle, \\ &=: S_n + J_n + E_n,\end{aligned}$$

where

$$\begin{aligned}S_n &= (nh)^{-2} \sum_{i_1,i_2=1}^{n} m\left(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}\right) \left\langle (\nabla\nabla^T K)\left(h^{-1}\mathbf{D}_{-\psi(x)}\left(\mathbf{p}(x) - \mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}\right)\right) \mathbf{v}_{\psi(x)}, \mathbf{v}_{\psi(x)} \right\rangle, \\ J_n &= (nh)^{-2} \sum_{i_1,i_2=1}^{n} j_\tau\left(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}\right) \left\langle (\nabla\nabla^T K)\left(h^{-1}\mathbf{D}_{-\psi(x)}\left(\mathbf{p}(x) - \mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}\right)\right) \mathbf{v}_{\psi(x)}, \mathbf{v}_{\psi(x)} \right\rangle, \\ E_n &= (nh)^{-2} \sum_{i_1,i_2=1}^{n} \varepsilon_{i_1,i_2} \left\langle (\nabla\nabla^T K)\left(h^{-1}\mathbf{D}_{-\psi(x)}\left(\mathbf{p}(x) - \mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}\right)\right) \mathbf{v}_{\psi(x)}, \mathbf{v}_{\psi(x)} \right\rangle.\end{aligned}$$

Note we dropped the inputs (x,w,ψ) for the latter terms for sake of convenience. By Lemma B.1 with $g_1(\mathbf{z}) = m(\mathbf{z})$, $g_2(\mathbf{z}) = g_3(\mathbf{z}) \equiv \mathbf{v}_{\psi(x)}$, $f(\mathbf{z}) = \nabla\nabla^T K(\mathbf{z})$ such that $r_1 = 2$, $r_2 = r_3 = 0$, $j = 1$, obtaining

$$\begin{aligned}S_n &= m(\mathbf{p}(x)) \int_{[-1,1]^2} \left\langle \nabla\nabla^T K(\mathbf{z}) \mathbf{v}_{\psi(x)}, \mathbf{v}_{\psi(x)} \right\rangle \mathrm{d}\mathbf{z} + O_{x\in I}(h) + O_{x\in I}\left((nh)^{-1}\right) \\ &= O_{x\in I}(h) + O_{x\in I}\left((nh)^{-1}\right),\end{aligned}$$

where the last equation is due to Lemma 2.21 (i). Applying Lemma B.2 with f, g_2 and g_3 as above and Lemma 2.21 (viii) yield

$$\begin{aligned}J_n &= \tau(x) \int_{[-1,1]\times[0,1]} \left\langle \nabla\nabla^T K(\mathbf{z}) \mathbf{v}_{\psi(x)}, \mathbf{v}_{\psi(x)} \right\rangle \mathrm{d}\mathbf{z} + O_{x\in I}(h) + O_{x\in I}\left((nh)^{-1}\right) \\ &= -\tau(x)\cos(\psi(x))^2 K_2^{(1)}(0) + O_{x\in I}(h) + O_{x\in I}\left((nh)^{-1}\right).\end{aligned}$$

Using Lemma B.3 with the same functions, we get immediately $E_n = o_{p,x\in I}(h)$. Hence,

$$\partial_w^2 \hat{\mathbb{M}}_n(0,\psi(x);x) = S_n + J_n + E_n \xrightarrow{\mathrm{P}} -\tau(x)\cos(\psi(x))^2 K_2^{(1)}(0).$$

As all the O-terms are uniform in x we have actually stochastic uniform convergence in the preceding display.

Step 2:

We show that
$$\sup_{x\in I} |\partial_w \partial_\psi \hat{\mathbb{M}}_n(0,\psi(x);x)| = O_P(h) + O_P((nh)^{-1}).$$

Use Lemma 2.19 and split $\partial_w \partial_\psi \hat{\mathbb{M}}_n(0,\psi(x);x)$ into six terms

$$\begin{aligned}
&\partial_w \partial_\psi \hat{\mathbb{M}}_n(0,\psi(x);x)\\
&\quad = \left(n^2h^3\right)^{-1} \sum_{i_1,i_2=1}^{n} Y_{i_1,i_2} \left\langle (\nabla\nabla^T K)\left(h^{-1}\mathbf{D}_{-\psi(x)}\left(\mathbf{p}(x)-\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}\right)\right)\mathbf{v}_{\psi(x)}, T\left(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}};0,\psi(x)\right)\right\rangle\\
&\qquad + \left(n^2h^2\right)^{-1} \sum_{i_1,i_2=1}^{n} Y_{i_1,i_2} \left\langle (\nabla K)\left(h^{-1}\mathbf{D}_{-\psi(x)}\left(\mathbf{p}(x)-\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}\right)\right)^T, (\cos(\psi(x)), -\sin(\psi(x)))\right\rangle,\\
&\quad =: S_{n,1} + J_{n,1} + E_{n,1} + S_{n,2} + J_{n,2} + E_{n,2},
\end{aligned}$$

where

$$\begin{aligned}
S_{n,1} &= \left(n^2h^3\right)^{-1} \sum_{i_1,i_2=1}^{n} m\left(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}\right) \left\langle (\nabla\nabla^T K)\left(h^{-1}\mathbf{D}_{-\psi(x)}\left(\mathbf{p}(x)-\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}\right)\right)\mathbf{v}_{\psi(x)}, T\left(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}};0,\psi(x)\right)\right\rangle,\\
J_{n,1} &= \left(n^2h^3\right)^{-1} \sum_{i_1,i_2=1}^{n} j_\tau\left(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}\right) \left\langle (\nabla\nabla^T K)\left(h^{-1}\mathbf{D}_{-\psi(x)}\left(\mathbf{p}(x)-\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}\right)\right)\mathbf{v}_{\psi(x)}, T\left(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}};0,\psi(x)\right)\right\rangle,\\
E_{n,1} &= \left(n^2h^3\right)^{-1} \sum_{i_1,i_2=1}^{n} \boldsymbol{\varepsilon}_{i_1,i_2} \left\langle (\nabla\nabla^T K)\left(h^{-1}\mathbf{D}_{-\psi(x)}\left(\mathbf{p}(x)-\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}\right)\right)\mathbf{v}_{\psi(x)}, T\left(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}};0,\psi(x)\right)\right\rangle,\\
S_{n,2} &= \left(n^2h^2\right)^{-1} \sum_{i_1,i_2=1}^{n} m\left(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}\right) \left\langle (\nabla K)\left(h^{-1}\mathbf{D}_{-\psi(x)}\left(\mathbf{p}(x)-\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}\right)\right)^T, (\cos(\psi(x)), -\sin(\psi(x)))\right\rangle,\\
J_{n,2} &= \left(n^2h^2\right)^{-1} \sum_{i_1,i_2=1}^{n} j_\tau\left(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}\right) \left\langle (\nabla K)\left(h^{-1}\mathbf{D}_{-\psi(x)}\left(\mathbf{p}(x)-\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}\right)\right)^T, (\cos(\psi(x)), -\sin(\psi(x)))\right\rangle,\\
E_{n,2} &= \left(n^2h^2\right)^{-1} \sum_{i_1,i_2=1}^{n} \boldsymbol{\varepsilon}_{i_1,i_2} \left\langle (\nabla K)\left(h^{-1}\mathbf{D}_{-\psi(x)}\left(\mathbf{p}(x)-\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}\right)\right)^T, (\cos(\psi(x)), -\sin(\psi(x)))\right\rangle.
\end{aligned}$$

By Lemma B.1 with $g_1(\mathbf{z}) = m(\mathbf{z})$, $f(\mathbf{z}) = \nabla\nabla^T K(\mathbf{z})$, $g_2 \equiv \mathbf{v}_{\psi(x)}$, $g_3(\mathbf{z}) = T(\mathbf{z};0,\psi(x))$, where $r_1 = 3$, $r_2 = 0$, $r_3 = j = 1$, together with (2.29) lead to

$$\begin{aligned}
S_{n,1} &= m(\mathbf{p}(x)) \int_{[-1,1]^2} \left\langle \nabla\nabla^T K(\mathbf{z})\mathbf{v}_{\psi(x)}, (z_2,-z_1)^T\right\rangle \mathrm{d}z_1\mathrm{d}z_2 + O_{x\in I}(h) + O_{x\in I}\left((nh)^{-1}\right)\\
&= O_{x\in I}(h) + O_{x\in I}\left((nh)^{-1}\right),
\end{aligned}$$

where the last line is due to Lemma 2.21 (ii) and (iii). Similarly, applying Lemma B.2 we get

$$\begin{aligned}
J_{n,1} &= \tau(x) \int_{[-1,1]\times[0,1]} \left\langle \nabla\nabla^T K(z_1,z_2)\mathbf{v}_{\psi(x)}, (z_2,-z_1)^T\right\rangle \mathrm{d}z_1\mathrm{d}z_2 + O_{x\in I}(h) + O_{x\in I}\left((nh)^{-1}\right)\\
&= O_{x\in I}(h) + O_{x\in I}\left((nh)^{-1}\right),
\end{aligned}$$

in which the second equality follows by Lemma 2.21 (ix) and (x). Using Lemma B.3 we get immediately $E_{n,1} = o_{p,x\in I}(h)$. Further, we use Lemma B.1 with $g_1(\mathbf{z}) = m(\mathbf{z})$, $f(\mathbf{z}) = (\nabla K(\mathbf{z})\ \nabla K(\mathbf{z}))$, $g_2 \equiv \mathbf{e}_1$ and $g_3(\mathbf{z}) = (\cos(\psi(x)), -\sin(\psi(x)))^T$, so that $r_1 = 2$, $r_2 = r_3 = 0$, $j = 1$, which together with

Lemma 2.21 (xi) implies

$$\begin{aligned}S_{n,2} &= m(\mathbf{p}(x)) \int_{[-1,1]^2} \langle \nabla K(\mathbf{z}), (\cos(\psi(x)), -\sin(\psi(x))) \rangle \, \mathrm{d}\mathbf{z} + O_{x\in I}(h) + O_{x\in I}\left((nh)^{-1}\right) \\ &= O_{x\in I}(h) + O_{x\in I}\left((nh)^{-1}\right).\end{aligned}$$

Using for Lemma B.2 the same functions as just and Lemma 2.21 (xi) one gets

$$\begin{aligned}J_{n,2} &= \tau(x) \int_{[-1,1]\times[0,1]} \langle \nabla K(\mathbf{z}), (\cos(\psi(x)), -\sin(\psi(x))) \rangle \, \mathrm{d}\mathbf{z} + O_{x\in I}(h) + O_{x\in I}\left((nh)^{-1}\right) \\ &= O_{x\in I}(h) + O_{x\in I}\left((nh)^{-1}\right).\end{aligned}$$

By means of Lemma B.3, obtain $E_{n,2} = o_{p,x\in I}(h)$. Summarizing, we deduce that

$$\partial_w \partial_\psi \hat{\mathbb{M}}_n(0,\psi(x);x) = S_{n,1} + J_{n,1} + E_{n,1} + S_{n,2} + J_{n,2} + E_{n,2} \xrightarrow{\mathrm{P}} 0.$$

As the O-terms are all uniform in x the latter display holds uniformly in x.

Step 3:

It remains to prove

$$\sup_{x\in I} |\partial_\psi^2 \hat{\mathbb{M}}_n(0,\psi(x);x) + \tau(x) K_2^{(1)}(0) \int_{-1}^{1} K_1(y) y^2 \, \mathrm{d}y| = O_P(h) + O_P((nh)^{-1}).$$

Use Lemma 2.19 and split $\partial_\psi^2 \hat{\mathbb{M}}_n(0,\psi(x);x)$ into six terms

$$\begin{aligned}&\partial_\psi^2 \hat{\mathbb{M}}_n(0,\psi(x);x) \\ &= \left(n^2h^4\right)^{-1} \sum_{i_1,i_2=1}^{n} Y_{i_1,i_2} \left\langle (\nabla\nabla^T K)\left(h^{-1}\mathbf{D}_{-\psi(x)}\left(\mathbf{p}(x)-\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}\right)\right) T(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}};0,\psi(x)), T(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}};0,\psi(x)) \right\rangle \\ &\quad - \left(n^2h^3\right)^{-1} \sum_{i_1,i_2=1}^{n} Y_{i_1,i_2} \left\langle (\nabla K)\left(h^{-1}\mathbf{D}_{-\psi(x)}\left(\mathbf{p}(x)-\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}\right)\right), \mathbf{D}_{-3\pi/2}\, T(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}};0,\psi(x)) \right\rangle \\ &=: S_{n,1} + J_{n,1} + E_{n,1} - S_{n,2} - J_{n,2} - E_{n,2},\end{aligned}$$

in which

$$\begin{aligned}S_{n,1} &= \left(n^2h^4\right)^{-1} \sum_{i_1,i_2=1}^{n} m\left(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}\right) \left\langle (\nabla\nabla^T K)\left(h^{-1}\mathbf{D}_{-\psi(x)}\left(\mathbf{p}(x)-\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}\right)\right) T(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}};0,\psi(x)), T(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}};0,\psi(x)) \right\rangle, \\ J_{n,1} &= \left(n^2h^4\right)^{-1} \sum_{i_1,i_2=1}^{n} j_\tau\left(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}\right) \left\langle (\nabla\nabla^T K)\left(h^{-1}\mathbf{D}_{-\psi(x)}\left(\mathbf{p}(x)-\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}\right)\right) T(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}};0,\psi(x)), T(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}};0,\psi(x)) \right\rangle, \\ E_{n,1} &= \left(n^2h^4\right)^{-1} \sum_{i_1,i_2=1}^{n} \varepsilon_{i_1,i_2} \left\langle (\nabla\nabla^T K)\left(h^{-1}\mathbf{D}_{-\psi(x)}\left(\mathbf{p}(x)-\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}\right)\right) T(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}};0,\psi(x)), T(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}};0,\psi(x)) \right\rangle, \\ S_{n,2} &= \left(n^2h^3\right)^{-1} \sum_{i_1,i_2=1}^{n} m\left(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}\right) \left\langle (\nabla K)\left(h^{-1}\mathbf{D}_{-\psi(x)}\left(\mathbf{p}(x)-\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}\right)\right), \mathbf{D}_{-3\pi/2}\, T(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}};0,\psi(x)) \right\rangle, \\ J_{n,2} &= \left(n^2h^3\right)^{-1} \sum_{i_1,i_2=1}^{n} j_\tau\left(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}\right) \left\langle (\nabla K)\left(h^{-1}\mathbf{D}_{-\psi(x)}\left(\mathbf{p}(x)-\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}\right)\right), \mathbf{D}_{-3\pi/2}\, T(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}};0,\psi(x)) \right\rangle, \\ E_{n,2} &= \left(n^2h^3\right)^{-1} \sum_{i_1,i_2=1}^{n} \varepsilon_{i_1,i_2} \left\langle (\nabla K)\left(h^{-1}\mathbf{D}_{-\psi(x)}\left(\mathbf{p}(x)-\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}\right)\right), \mathbf{D}_{-3\pi/2}\, T(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}};0,\psi(x)) \right\rangle.\end{aligned}$$

Application of Lemma B.1 with the functions $g_1(\mathbf{z}) = m(\mathbf{z})$, $f(\mathbf{z}) = \nabla\nabla^T K(\mathbf{z})$, $g_2(\mathbf{z}) = g_3(\mathbf{z}) = T(\mathbf{z};0,\psi(x))$ ($r_1 = 4$, $r_2 = r_3 = 1$, $j = 1$) yields with (2.29)

$$
\begin{aligned}
S_{n,1} &= m(\mathbf{p}(x)) \int_{[-1,1]^2} \left\langle \nabla\nabla^T K(z_1,z_2)(z_2,-z_1)^T, (z_2,-z_1)^T \right\rangle \mathrm{d}z_1 \mathrm{d}z_2 + O_{x\in I}(h) + O_{x\in I}\left((nh)^{-1}\right) \\
&= O_{x\in I}(h) + O_{x\in I}\left((nh)^{-1}\right),
\end{aligned}
$$

where the second equality is by Lemma 2.21 (v) and (vi). With g_2 and f as just and Lemma B.2 combined with Lemma 2.21 (xii) and (xiii),

$$
\begin{aligned}
J_{n,1} &= \tau(x) \int_{[-1,1]\times[0,1]} \left\langle \nabla\nabla^T K(z_1,z_2)(z_2,-z_1)^T, (z_2,-z_1)^T \right\rangle \mathrm{d}z_1 \mathrm{d}z_2 + O_{x\in I}(h) + O_{x\in I}\left((nh)^{-1}\right) \\
&= -2\tau(x) - \tau(x) K_2^{(1)}(0) \int_{-1}^{1} K_1(y)\, y^2 \,\mathrm{d}y + O_{x\in I}(h) + O_{x\in I}\left((nh)^{-1}\right).
\end{aligned}
$$

By Lemma B.3 obtain that $E_{n,1} = o_{p,x\in I}(h)$. Next, using $g_1(\mathbf{z}) = m(\mathbf{z})$, $f(\mathbf{z}) = (\nabla K(\mathbf{z})\ \nabla K(\mathbf{z}))$, $g_2 \equiv \mathbf{e}_1$ and $g_3(\mathbf{z}) = \mathbf{D}_{-3\pi/2} T(\mathbf{z};0,\psi(x))$ in Lemma B.1 ($r_1 = 3$, $r_2 = 0$, $r_3 = j = 1$) and Lemma 2.21 (vii) yield

$$
\begin{aligned}
S_{n,2} &= m(\mathbf{p}(x)) \int_{[-1,1]^2} \left\langle \nabla K(z_1,z_2), (z_1,z_2)^T \right\rangle \mathrm{d}z_1 \mathrm{d}z_2 + O_{x\in I}(h) + O_{x\in I}\left((nh)^{-1}\right) \\
&= O_{x\in I}(h) + O_{x\in I}\left((nh)^{-1}\right).
\end{aligned}
$$

With g_2, g_3 and f as just, we get by applying Lemma B.2 and Lemma 2.21 (xiv),

$$
\begin{aligned}
J_{n,2} &= \tau(x) \int_{[-1,1]\times[0,1]} \left\langle \nabla K\left((z_1,z_2)^T\right), (z_1,z_2)^T \right\rangle \mathrm{d}z_1 \mathrm{d}z_2 + O_{x\in I}(h) + O_{x\in I}\left((nh)^{-1}\right) \\
&= -2\tau(x) + O_{x\in I}(h) + O_{x\in I}\left((nh)^{-1}\right).
\end{aligned}
$$

and $E_{n,2} = o_{p,x\in I}(h)$ from Lemma B.3. Finally, one obtains uniformly in x that

$$
\partial_\psi^2 \hat{\mathbb{M}}_n(0,\psi(x);x) = S_{n,1} + J_{n,1} + E_{n,1} - S_{n,2} - J_{n,2} - E_{n,2} \xrightarrow{\mathrm{P}} -\tau(x) K_2^{(1)}(0) \int_{-1}^{1} K_1(y)\, y^2 \,\mathrm{d}y.
$$

Lipschitz continuity

It suffices to show that

$$
\| \nabla\nabla^T \hat{\mathbb{M}}_n(w_1,\psi_1;x) - \nabla\nabla^T \hat{\mathbb{M}}_n(w_2,\psi_2;x) \| \le L_n \| (w_1,\psi_1)^T - (w_2,\psi_2)^T \|, \tag{2.31}
$$

where $L_n = O_{p,x\in I}(1)$ and $(w_i,\psi_i) \in \Theta_n$, $i = 1,2$. By taking

$$
L_n = \sup_{(x,w,\psi)^T \in \tilde{\Theta}_{n,x}} \| (\nabla) \otimes (\nabla\nabla^T) \hat{\mathbb{M}}_n(w,\psi;x) \|,
$$

where $\otimes$ is the Kronecker product, we may obtain (2.31) by the mean value theorem. One can show that all components of $(\nabla)\otimes(\nabla\nabla^T)\hat{\mathbb{M}}_n(w,\psi;x)$ are uniformly bounded in probability. For example, letting as above $\mathbf{v}_\psi = (\sin(\psi),\cos(\psi))$, then the third partial derivative for w is

$$
\begin{aligned}
&\partial_w^3 \hat{\mathbb{M}}_n(w,\psi;x) \\
&= (nh)^{-2} \sum_{i_1,i_2=1}^{n} Y_{i_1,i_2} \left\langle (\mathbf{v}_\psi^T \otimes \nabla) \otimes (\nabla\nabla^T K)(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}(x) + wh\mathbf{e_2} - \mathbf{x_{i_1,i_2}}))(\mathbf{v}_\psi,\mathbf{v}_\psi)^T, \mathbf{v}_\psi^T \right\rangle.
\end{aligned}
$$

Let $I(n,h)$ denote the set of indices for which the latter sum is not zero. Hence, $|I(n,h)| \le 4n^2h^2$ and therefore

$$\mathbb{E} \sup_{(x,w,\psi)^T \in \Theta_n} |\partial_w^3 \hat{\mathbb{M}}_n(w,\psi;x)| \le 4(nh)^{-2} ||(\nabla \otimes \nabla\nabla^T)K||_\infty \sum_{i_1,i_2 \in I(n,h)} \mathbb{E}|Y_{i_1,i_2}| = O(1).$$

The remaining partial derivatives are dealt with similarly. □

Proof of Lemma 2.11. Note that $\nabla\nabla^T \mathbb{M}_n(0,\psi(x);x)$ has the same components as $\nabla\nabla^T \hat{\mathbb{M}}_n(0,\psi(x);x)$ besides the stochastic parts, which were denoted by E_n in the proof of Lemma 2.21 and which were stochastically negligible. Hence, $\nabla\nabla^T \mathbb{M}_n(0,\psi(x);x)$ has the same limit as $\nabla\nabla^T \hat{\mathbb{M}}_n(0,\psi(x);x)$ and this implies the convergence of the deterministic Hessian matrix $\mathbf{H}_n(x)$ against the same limit as the stochastic Hessian matrix $\hat{\mathbf{H}}_n(x)$.

□

2.6.3 Asymptotic normality of the score: proof of Lemma 2.10

Proof of Lemma 2.10. By means of Lemma 2.17 and the definition of $\mathbb{M}_n$, it is apparently that

$$\nabla \mathbb{M}_n(w,\psi;x) = \mathbb{E}\big(\nabla \hat{\mathbb{M}}_n(w,\psi;x)\big).$$

Let $x \in I$. We intend to make use of the Lindeberg-Feller Theorem (see, e.g. van der Vaart (2000), Proposition 2.27). By Lemma 2.17

$$nh\Big(\nabla \hat{\mathbb{M}}_n(0,\psi(x);x) - \mathbb{E}\big(\nabla \hat{\mathbb{M}}_n(0,\psi(x);x)\big)\Big) = \sum_{i_1,i_2=1}^{n} \mathbf{a}_{\mathbf{i_1},\mathbf{i_2}} \varepsilon_{i_1,i_2},$$

with

$$\begin{aligned}
\left(\mathbf{a}_{\mathbf{i_1},\mathbf{i_2}}\right)_1 &= (nh)^{-1} \Big\langle (\nabla K)\Big(h^{-1}\mathbf{D}_{-\psi(x)}\left(\mathbf{p}(x) - \mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}\right)\Big), (\sin(\psi(x)), \cos(\psi(x)))^T \Big\rangle,\\
\left(\mathbf{a}_{\mathbf{i_1},\mathbf{i_2}}\right)_2 &= \left(nh^2\right)^{-1} \Big\langle (\nabla K)\Big(h^{-1}\mathbf{D}_{-\psi(x)}\left(\mathbf{p}(x) - \mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}\right)\Big), T\left(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}; 0, \psi(x)\right) \Big\rangle.
\end{aligned}$$

First, we consider the convergence of the covariance matrix,

$$\sum_{i_1,i_2=1}^{n} \mathrm{Cov}\left(\mathbf{a}_{\mathbf{i_1},\mathbf{i_2}} \varepsilon_{i_1,i_2}\right) = \sigma^2 \sum_{i_1,i_2=1}^{n} \begin{pmatrix} \left(\mathbf{a}_{\mathbf{i_1},\mathbf{i_2}}\right)_1^2 & \left(\mathbf{a}_{\mathbf{i_1},\mathbf{i_2}}\right)_1 \left(\mathbf{a}_{\mathbf{i_1},\mathbf{i_2}}\right)_2 \\ \left(\mathbf{a}_{\mathbf{i_1},\mathbf{i_2}}\right)_1 \left(\mathbf{a}_{\mathbf{i_1},\mathbf{i_2}}\right)_2 & \left(\mathbf{a}_{\mathbf{i_1},\mathbf{i_2}}\right)_2^2 \end{pmatrix} \to \sigma^2 \mathrm{diag}(V_N(x), W_N).$$

Notice that we suppress the dependence of $\mathbf{a}_{\mathbf{i_1},\mathbf{i_2}}$ on x in the notation.

Step 1:

We show

$$\sum_{i_1,i_2=1}^{n} \left(\mathbf{a}_{\mathbf{i_1},\mathbf{i_2}}\right)_1^2 \to V_N(x).$$

Applying Lemma B.1 with $g_1 \equiv 1$, $f(\mathbf{z}) = (\nabla K(\mathbf{z})\ \nabla K(\mathbf{z}))$, $g_2 \equiv \mathbf{e}_1$, $g_3 \equiv (\sin\psi(x), \cos\psi(x))^T$, such that $r_1 = 2$, $r_2 = r_3 = 0$, and $j = 2$, we obtain

$$\sum_{i_1,i_2=1}^{n} \left(\mathbf{a}_{\mathbf{i_1},\mathbf{i_2}}\right)_1^2 = \int_{[-1,1]^2} \left\langle \nabla K(\mathbf{z}), (\sin\psi(x), \cos\psi(x))^T \right\rangle^2 \mathrm{d}\mathbf{z} + O(h) + O\left((nh)^{-1}\right)$$

$$
\begin{aligned}
&= \sin^2(\psi(x)) \int_{[-1,1]^2} K^{(1,0)}(\mathbf{z})^2 \,\mathrm{d}\mathbf{z} + 2\cos(\psi(x))\sin(\psi(x)) \int_{[-1,1]^2} K^{(1,0)}(\mathbf{z}) K^{(0,1)}(\mathbf{z}) \,\mathrm{d}\mathbf{z} \\
&\quad + \cos^2(\psi(x)) \int_{[-1,1]^2} K^{(0,1)}(\mathbf{z})^2 \,\mathrm{d}\mathbf{z} + O(h) + O\left((nh)^{-1}\right) \\
&= V_N(x) + O(h) + O\left((nh)^{-1}\right),
\end{aligned}
$$

since $y \mapsto K_i^{(1)}(y) K_i(y)$ are odd functions, $i = 1,2$.

Step 2:

We show

$$
\sum_{i_1,i_2=1}^{n} \left(\mathbf{a}_{\mathbf{i}_1,\mathbf{i}_2}\right)_2^2 \to W_N.
$$

Applying Lemma B.1 with $g_1 \equiv 1$, $f(\mathbf{z}) = (\nabla K(\mathbf{z})\ \nabla K(\mathbf{z}))$, $g_2 \equiv \mathbf{e}_1$, $g_3(\mathbf{z}) = T(\mathbf{z}; 0, \psi(x))$, such that $r_1 = 4$, $r_2 = 0$, $r_3 = 1$, $j = 2$, and using (2.29) lead to

$$
\begin{aligned}
\sum_{i_1,i_2=1}^{n} \left(\mathbf{a}_{\mathbf{i}_1,\mathbf{i}_2}\right)_2^2 &= \int_{[-1,1]^2} \left\langle \nabla K(z_1,z_2), (z_2,-z_1)^T \right\rangle^2 \mathrm{d}z_1 \mathrm{d}z_2 + O(h) + O\left(\left(nh^2\right)^{-1}\right) \\
&= W_N + O(h) + O\left(\left(nh^2\right)^{-1}\right).
\end{aligned}
$$

Step 3:

We show

$$
\sum_{i_1,i_2=1}^{n} \left(\mathbf{a}_{\mathbf{i}_1,\mathbf{i}_2}\right)_1 \left(\mathbf{a}_{\mathbf{i}_1,\mathbf{i}_2}\right)_2 \to 0.
$$

Applying Lemma B.1 with $g_1 \equiv 1$, f and g_2 as before, $g_3(\mathbf{z}) = (\sin\psi(x), \cos\psi(x))^T$, $g_4(\mathbf{z}) = T(\mathbf{z}; 0, \psi(x))$, such that $r_1 = 3$, $r_2 = r_3 = 0$, $r_4 = 1$, and using (2.29)

$$
\begin{aligned}
&\sum_{i_1,i_2=1}^{n} \left(\mathbf{a}_{\mathbf{i}_1,\mathbf{i}_2}\right)_1 \left(\mathbf{a}_{\mathbf{i}_1,\mathbf{i}_2}\right)_2 \\
&\quad = \int_{[-1,1]^2} \left\langle \nabla K(z_1,z_2), (\sin\psi(x), \cos\psi(x))^T \right\rangle \left\langle \nabla K(z_1,z_2), (z_2,-z_1)^T \right\rangle \mathrm{d}z_1 \mathrm{d}z_2 \\
&\qquad + O(h) + O\left(\left(nh^2\right)^{-1}\right) \\
&\quad = \sin(\psi(x)) \int_{[-1,1]^2} K^{(1,0)}(z_1,z_2)^2 z_2 \,\mathrm{d}z_1 \mathrm{d}z_2 \\
&\qquad - \sin(\psi(x)) \int_{[-1,1]^2} K^{(1,0)}(z_1,z_2) K^{(0,1)}(z_1,z_2) z_1 \,\mathrm{d}z_1 \mathrm{d}z_2 \\
&\qquad + \cos(\psi(x)) \int_{[-1,1]^2} K^{(1,0)}(z_1,z_2) K^{(0,1)}(z_1,z_2) z_2 \,\mathrm{d}z_1 \mathrm{d}z_2 \\
&\qquad - \cos(\psi(x)) \int_{[-1,1]^2} K^{(0,1)}(z_1,z_2)^2 z_1 \,\mathrm{d}z_1 \mathrm{d}z_2 + O(h) + O\left(\left(nh^2\right)^{-1}\right) \\
&\quad = O(h) + O\left(\left(nh^2\right)^{-1}\right),
\end{aligned}
$$

where the last line follows by the fact that $y \mapsto K_i^2(y)y$, $i = 1,2$, are odd, so the first and fourth integral vanish, and $y \mapsto K_i^{(1)}(y)K_i(y)$, $i = 1,2$, are odd as well, so that the second and third integral also vanish.

Step 4:

We check the Lindeberg condition: For any $\epsilon > 0$

$$\sum_{i_1,i_2=1}^{n} ||\mathbf{a}_{\mathbf{i}_1,\mathbf{i}_2}||_2^2 \,\mathbb{E}\Big(\varepsilon_{i_1,i_2}^2 1_{\{\,||\mathbf{a}_{\mathbf{i}_1,\mathbf{i}_2}||_2\ |\varepsilon_{i_1,i_2}|>\epsilon\}}\Big) = o(1).$$

Notice that

$$\begin{aligned}\max_{1\le i_1,i_2\le n} ||\mathbf{a}_{\mathbf{i}_1,\mathbf{i}_2}||_2^2 &\le \max_{1\le i_1,i_2\le n} \big(\mathbf{a}_{\mathbf{i}_1,\mathbf{i}_2}\big)_1^2 + \max_{1\le i_1,i_2\le n} \big(\mathbf{a}_{\mathbf{i}_1,\mathbf{i}_2}\big)_2^2 \\ &\le 2||\nabla K||_\infty \Big(\big(n^2h^2\big)^{-1} + \big(n^2h^4\big)^{-1}\Big) = o(1),\end{aligned} \tag{2.32}$$

and by Step 1 and Step 2

$$\sum_{i_1,i_2=1}^{n} ||\mathbf{a}_{\mathbf{i}_1,\mathbf{i}_2}||_2^2 = \sum_{i_1,i_2=1}^{n} \big(\mathbf{a}_{\mathbf{i}_1,\mathbf{i}_2}\big)_1^2 + \sum_{i_1,i_2=1}^{n} \big(\mathbf{a}_{\mathbf{i}_1,\mathbf{i}_2}\big)_2^2 \to V_N(x) + W_N. \tag{2.33}$$

Thus, by (2.32) and (2.33)

$$\sum_{i_1,i_2=1}^{n} ||\mathbf{a}_{\mathbf{i}_1,\mathbf{i}_2}||_2^2 \,\mathbb{E}\Big(\varepsilon_{i_1,i_2}^2 1_{\{\,||\mathbf{a}_{\mathbf{i}_1,\mathbf{i}_2}||_2\ |\varepsilon_{i_1,i_2}|>\epsilon\}}\Big) \le \max_{1\le i_1,i_2\le n} \mathbb{E}\Big(\varepsilon_{i_1,i_2}^2 1_{\{\,||\mathbf{a}_{\mathbf{i}_1,\mathbf{i}_2}||_2\ |\varepsilon_{i_1,i_2}|>\epsilon\}}\Big) \sum_{i_1,i_2=1}^{n} ||\mathbf{a}_{\mathbf{i}_1,\mathbf{i}_2}||_2^2 = o(1).$$

Eventually, the Lindeberg-Feller theorem concludes the lemma. □

2.6.4 Gaussian approximation: proofs of Lemmas 2.13, 2.15 and of Proposition 2.9

Proof of Lemma 2.13. By means of Lemma A.1 with

$$g_1(\mathbf{z}) \equiv V_N^{-1/2}(x), \quad g_2 \equiv \mathbf{e}_1, \quad g_3(\mathbf{z}) \equiv (\sin\psi, \cos\psi)^T, \quad f(\mathbf{z}) = (\nabla K\ \nabla K),$$

and $w = 0$ and $\psi = \psi(x)$, such that $r_1 = 1$ and $r_2 = r_3 = 0$, gives us that for n large enough

$$||Z_n^\phi - Z_{n,G}^\phi||_\infty = O_P\Big(\frac{\sqrt{\log(n)}}{n^{1/2}h}\Big).$$

The second part of Lemma A.3 with the same g_1, g_2, g_3 and f shows that $\mathbb{E}(\mathbf{M}_\phi) \le C\sqrt{\log(n)}$, for some constant $C > 0$ and $\mathbf{M}_\phi$ is given in (2.23). This implies by Markov's inequality that

$$\mathbf{M}_\phi = O_P\big(\sqrt{\log(n)}\big).$$

Theorem 2.1 in Chernozhukov et al. (2014) states that

$$\mathbb{L}\Big(\mathbf{M}_\phi, \delta_n\Big) \le 4\delta_n\Big(\mathbb{E}(\mathbf{M}_\phi) + 1\Big).$$

With the bound on the expected value and since $\delta_n = o(\log(n)^{-1/2})$ the right-hand side of the preceding display is $o(1)$. □

Proof of Lemma 2.15. Using Lemma A.1 and Lemma A.3 this time with

$$g_1(\mathbf{z}) \equiv W_N^{-1/2}, \quad g_2(\mathbf{z}) \equiv \mathbf{e}_1, \quad g_3(\mathbf{z}) = T(\mathbf{z}; 0, \psi(x)), \quad f(\mathbf{z}) = (\nabla K \ \nabla K),$$

and $w = 0, \psi = \psi(x)$ such that $r_1 = 2, r_2 = 0$ and $r_3 = 1$ yields the assertion analogously as in the proof of Lemma 2.13. □

Proof of Proposition 2.9. From (2.22)

$$n\,h\left(\partial_w \hat{\mathbb{M}}_n(0,\psi(x);x) - \partial_w \mathbb{M}_n(0,\psi(x);x)\right) = \sigma\, V_N(x)^{1/2} Z_n^{\phi}(x).$$

Further, $V_N(\cdot)$ is uniformly bounded over I. Hence, Lemma 2.13 implies

$$\begin{aligned}\sup_{x\in I}|n\,h\left(\partial_w \hat{\mathbb{M}}_n(0,\psi(x);x) - \partial_w \mathbb{M}_n(0,\psi(x);x)\right)| &= \sigma \sup_{x\in I}|\ V_N(x)^{1/2} Z_{n,G}^{\phi}(x)| + O_P\left(\frac{\sqrt{\log(n)}}{n^{1/2}h}\right)\\ &= O_P\left(\sqrt{\log(n)}\right) + o_P(1),\end{aligned}$$

which shows the assertion for the first component of

$$n\,h\left(\nabla \hat{\mathbb{M}}_n(0,\psi(x);x) - \nabla \mathbb{M}_n(0,\psi(x);x)\right).$$

The proof for the second component, that is $n\,h\left(\partial_\psi \hat{\mathbb{M}}_n(0,\psi(x);x) - \partial_\psi \mathbb{M}_n(0,\psi(x);x)\right)$, is analogously by using Lemma 2.15. □

Note that the statement of Proposition 2.9 for the estimates is stronger than Proposition 2.4. But the uniform consistency of the estimates is needed for the proof of Lemma 2.8, which is however crucial for the proof of Proposition 2.9.

2.6.5 Rate of convergence of the Hessian matrix: proof of Lemma 2.12

Lemma 2.22. *Let B be a compact subset of $\mathbb{R}$ and $\hat{A}_n, \hat{A} : B \to \mathbb{R}^{2\times 2}$ be matrix valued (random) functions with (stochastically) bounded marginals and*

$$||\,\hat{A}_n - \hat{A}\,||_\infty = O_P(r_n),$$

for some real-valued sequence r_n. If $\hat{A}_n^{-1}(x)$ and $\hat{A}^{-1}(x)$ exist almost surely for every $x \in B$, then

$$||\,\hat{A}_n^{-1} - \hat{A}^{-1}\,||_\infty = O_P(r_n).$$

Proof. First, note that for any $x \in B$ one has almost surely

$$||\,\hat{A}_n^{-1}(x) - \hat{A}^{-1}(x)\,|| \le ||\,\hat{A}^{-1}(x)\,|| \cdot ||\,\hat{A}_n(x) - \hat{A}(x)\,|| \cdot ||\,\hat{A}_n^{-1}(x)\,||.$$

Since $\hat{A}_n$ and A are stochastically bounded the inverse functions $\hat{A}_n^{-1}, A^{-1}$ are almost surely stochastically bounded. Therefore,

$$||\,\hat{A}_n^{-1} - \hat{A}^{-1}\,||_\infty \le ||\,\hat{A}^{-1}\,||_\infty ||\,\hat{A}_n - \hat{A}\,||_\infty ||\,\hat{A}_n^{-1}\,||_\infty = O_P(r_n).$$

□

Proof of Lemma 2.12. By compactness of I and Lipschitz continuity of cosine and sine it easily follows that

$$||\cos^2\circ\hat{\psi}_n - \cos^2\circ\psi||_\infty = O(||\hat{\psi}_n-\psi||_\infty), \quad ||\sin^2\circ\hat{\psi}_n - \sin^2\circ\psi||_\infty = O(||\hat{\psi}_n-\psi||_\infty),$$

so that by means of Proposition 2.4

$$\begin{aligned}||\hat{V}_H - V_H||_\infty &= K_2^{(1)}(0)\sup_{x\in I}|\hat{\tau}_n(x)\cos^2(\hat{\psi}_n(x)) - \tau(x)\cos^2(\psi(x))|\\ &\le O(||\hat{\psi}_n-\psi||_\infty) + O(||\hat{\tau}_n-\tau||_\infty)\\ &= O(||\hat{\psi}_n-\psi||_\infty) + O(||\hat{w}_n||_\infty) + O_P(h) + O((nh)^{-1}).\end{aligned}$$

By Proposition 2.9 the preceding display is $O_P(h)+O_P\big(\sqrt{\log(n)}/nh\big)$. Similarly,

$$||\hat{V}_N - V_N||_\infty = O_P\big(\sqrt{\log(n)}/nh\big), \quad ||\hat{W}_H - W_H||_\infty = O_P(h)+O_P\big(\sqrt{\log(n)}/nh\big).$$

In the proof of Lemma 2.8 we have shown that for $\tilde{\mathbf{H}}_n(x) = \nabla\nabla^T\hat{\mathbb{M}}_n(0,\psi(x);x)$

$$||\tilde{\mathbf{H}}_n - \mathbf{H}||_\infty = O_P(h) + O_P((nh)^{-1}).$$

In addition, by the stochastic Lipschitz continuity of $\nabla\nabla^T\hat{\mathbb{M}}_n$ obtain

$$||\hat{\mathbf{H}}_n - \tilde{\mathbf{H}}_n||_\infty \le O_P(1)\,||(\hat{w}_n(\cdot),(\hat{\psi}_n(x)-\psi)(\cdot))^T||_\infty = O_P\big(\sqrt{\log(n)}/nh\big),$$

by means of Theorem 2.1. Combine the last two displays to get $||\hat{\mathbf{H}}_n - \mathbf{H}_n||_\infty = O_P\big(\sqrt{\log(n)}/nh\big)$. Application of Lemma 2.22 completes the proof. □

2.6.6 Quantile approximation: proofs of Lemma 2.14 and Lemma 2.16

For the proofs we will need the following lemma.

Lemma 2.23. *Let $(X_n)_{n\in\mathbb{N}}$ and $(Y_n)_{n\in\mathbb{N}}$ be stochastic processes and each Y_n has a continuous distribution. Assume that $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ are sequences of positive real numbers with*

1. $b_n = o(1)$ *and* $a_n/b_n = o(1)$;
2. $|X_n - Y_n| = O_P(a_n)$ $n\longrightarrow\infty$;
3. *There exists a* $\zeta_0 > 0$ *with* $\limsup_{n\to\infty}\mathbb{L}(Y_n,\zeta b_n) \le \delta(\zeta)$ *for any* $\zeta\in(0,\delta(\zeta))$, *where* $\delta(\zeta)>0$ *and* $\lim_{\zeta\to 0}\delta(\zeta) = 0$.

Then,

$$\lim_{n\to\infty} P(X_n \le q_\alpha(Y_n)) = \alpha.$$

Proof of Lemma 2.23. First, it holds that

$$\limsup_n P(X_n \le q_\alpha(Y_n)) = \alpha.$$

Indeed, let $\zeta\in(0,\zeta_0)$, then

$$\begin{aligned}P(X_n \le q_\alpha(Y_n)) &\le P(Y_n \le q_\alpha(Y_n) + b_n\zeta) + P(|Y_n - X_n| \ge b_n\zeta)\\ &\le \alpha + P(q_\alpha(Y_n) \le Y_n \le q_\alpha(Y_n) + b_n\zeta) + P(|Y_n - X_n| \ge b_n\zeta)\end{aligned}$$

$$\leq \alpha + \mathbb{L}(Y_n, b_n\zeta) + P(|Y_n - X_n| \geq b_n\zeta).$$

By considering $n \to \infty$ and then $\zeta \searrow 0$ yields the assertion above. Next,

$$\liminf_n P(X_n \leq q_\alpha(Y_n)) = \alpha,$$

which would complete the proof. Analogously as just,

$$\begin{aligned} P(X_n \leq q_\alpha(Y_n)) &\geq P(Y_n \leq q_\alpha(Y_n) - b_n\zeta) - P(|Y_n - X_n| \geq b_n\zeta) \\ &\geq \alpha - P(q_\alpha(Y_n) - b_n\zeta \leq Y_n \leq q_\alpha(Y_n)) - P(|Y_n - X_n| \geq b_n\zeta) \\ &\geq \alpha - \mathbb{L}(Y_n, b_n\zeta) - P(|Y_n - X_n| \geq b_n\zeta). \end{aligned}$$

Again, considering $n \to \infty$ and then $\zeta \searrow 0$ leads to the claim. □

Proof of Lemma 2.14. We intend to make use of Lemma 2.23. For this purpose we break the proof down into two steps. In the first step, we verify

$$|\tilde{\mathbf{M}}_\phi - \mathbf{M}_\phi| = O_P\Big(\frac{\log(n)}{nh}\Big), \tag{2.34}$$

where $\tilde{\mathbf{M}}_\phi$ is given in (2.11) and $\mathbf{M}_\phi$ is given in (2.23). In the second step, it will be shown that

$$\lim_{\zeta \to 0} \limsup_{n \to \infty} \mathbb{L}(\tilde{\mathbf{M}}_\phi, \zeta b_n) = 0, \tag{2.35}$$

for some real-valued sequence $b_n = o(1)$ such that ${}^{\log(n)}/_{b_n nh} = o(1)$.

Step 1: Verifying (2.34)

Using Lemma A.1 with ξ_{i_1,i_2} instead of $\boldsymbol{\varepsilon}_{i_1,i_2}$ which can be done as ξ_{i_1,i_2} has the same independence properties as $\boldsymbol{\varepsilon}_{i_1,i_2}$ but stricter moment properties. Setting in that context $\sigma = 1$, $g_1(\mathbf{z}) \equiv V_N^{-1/2}(x)$, $g_2 \equiv \mathbf{e}_1$, $g_3(\mathbf{z}) \equiv (\sin\psi, \cos\psi)^T$, $f(\mathbf{z}) = (\nabla K \ \nabla K)$, $w = \hat{w}_n(x)$ and $\psi = \hat{\psi}_n(x)$, such that $r_1 = 1$ and $r_2 = r_3 = 0$ gives us that

$$||\tilde{Z}_n^\phi - \tilde{Z}_{n,G}^\phi||_\infty = O_P\Big(\frac{\sqrt{\log(n)}}{n^{1/2}h}\Big), \tag{2.36}$$

where

$$\tilde{Z}_{n,G}^\phi(x) = \frac{1}{\sqrt{\hat{V}_N(x)}} \int_{\mathbb{R}^2} \Big\langle (\nabla K)\Big(\mathbf{D}_{-\hat{\psi}_n(x)}(h^{-1}(x, \hat{\phi}_n(x))^T - \mathbf{z})\Big), (\sin\hat{\psi}_n(x), \cos\hat{\psi}_n(x))^T \Big\rangle \, \mathrm{d}W(\mathbf{z}).$$

For sake of brevity write

$$F(D_{-\psi}(\mathbf{p}_{w,h}(x)/h - \mathbf{z}); w, \psi) = g_1(h\mathbf{z}) \big\langle f(\mathbf{D}_{-\psi}(\mathbf{p}(x)/h + w\mathbf{e}_2 - \mathbf{z}))g_2(h\mathbf{z}), g_3(h\mathbf{z}) \big\rangle$$

and note that we suppressed the dependency of g_i for $i = 1,2,3$ on w and ψ in the notation. Moreover,

$$\begin{aligned} \tilde{Z}_{n,G}^\phi(x) &= \int_{\mathbb{R}^2} F(D_{-\hat{\psi}_n(x)}(h^{-1}(x, \hat{\phi}_n(x))^T - \mathbf{z}); \hat{w}_n(x), \hat{\psi}_n(x)) \, \mathrm{d}W(\mathbf{z}) \\ Z_{n,G}^\phi(x) &= \int_{\mathbb{R}^2} F(D_{-\psi}(\mathbf{p}(x)/h - \mathbf{z}); 0, \psi(x)) \, \mathrm{d}W(\mathbf{z}). \end{aligned}$$

Choosing $\delta = O_P\big(\sqrt{\log(n)}/nh\big)$, then the third part of Lemma A.3 together with Theorem 2.1 implies

$$||\,\tilde{Z}^{\phi}_{n,G} - Z^{\phi}_{n,G}\,||_\infty = O_P\big(\log(n)/nh^2\big).$$

With this and (2.36) we have that

$$||\,\tilde{Z}^{\phi}_{n} - Z^{\phi}_{n,G}\,||_\infty \le ||\,\tilde{Z}^{\phi}_{n} - \tilde{Z}^{\phi}_{n,G}\,||_\infty + ||\,\tilde{Z}^{\phi}_{n,G} - Z^{\phi}_{n,G}\,||_\infty = O_P\Big(\frac{\sqrt{\log(n)}}{n^{1/2}h}\Big) + O_P\Big(\frac{\log(n)}{nh^2}\Big)$$

which implies (2.34) with the triangle inequality.

Step 2: Verifying (2.35)

For any $\zeta > 0$

$$\begin{aligned}\mathbb{L}(\tilde{\mathbf{M}}_\phi, \zeta b_n) &\le P\big(|\tilde{\mathbf{M}}_\phi - \mathbf{M}_\phi| > \zeta b_n\big) + \sup_{x\in I} P\big(|\tilde{\mathbf{M}}_\phi - x| \le \zeta b_n, |\tilde{\mathbf{M}}_\phi - \mathbf{M}_\phi| \le \zeta b_n\big)\\ &\le P\big(|\tilde{\mathbf{M}}_\phi - \mathbf{M}_\phi| > \zeta b_n\big) + \mathbb{L}(\mathbf{M}_\phi, 2\zeta b_n).\end{aligned}$$

By means of Lemma 2.13, if $b_n = o(\log(n)^{-1/2})$ the Lévy concentration function in the preceding display is $o(1)$. In view of (2.34) the first summand on the right-hand side of the latter inequality can be made arbitrary small provided that $\frac{\log(n)^{3/2}}{nh} = o(1)$, which is implied by Assumption 2.4. Hence, we obtain (2.35).

Step 3: Asymptotic behavior of the quantile

The fourth part of Lemma A.3 states that $q_{1-\alpha}(\mathbf{M}_\phi) \cong \sqrt{\log(n)}$. Due to Lemma 2.23 the quantile $q_{1-\alpha}(\tilde{\mathbf{M}}_\phi)$ must be of the same order as $q_{1-\alpha}(\mathbf{M}_\phi)$. □

Proof of Lemma 2.16. The proof follows exactly the same ideas as the proof before by using the corresponding results for the score process in (2.11), (2.24) and (2.25) of the jump-slope. □

CHAPTER 3

Adaptive confidence intervals for kink estimation

In this chapter we construct asymptotic confidence sets for the location and the magnitude of the jump in the γ-th derivative (kink) of a univariate regression curve, which is assumed to have at least $s \geq \gamma + 1$ continuous derivatives outside the kink as well as a suitably smooth extension of the $\gamma + 1$-th derivative at the kink. To this end, the asymptotic normal distribution of the zero-crossing-time estimator by Goldenshluger et al. (2006) is derived based on methods from Z-estimation. Through a Lepski-choice of the tuning parameter, the resulting confidence sets are adaptive with respect to s over smaller, separated function classes which allow for an explicit control of the bias term.

This chapter is structured as follows. In Section 3.1 the model is introduced as well as the estimates based on the zero-crossing-time-technique. Results on their rate of convergence are given in Section 3.1 as well, while in Section 3.2 the joint asymptotic normality of the estimates is stated and afterwards adaptive confidence intervals are constructed. The proofs are given in Section 3.5. The finite-sample performance of the suggested adaptive confidence interval for the kink-location is investigated in a simulation study for artificial data as well as for real-world datasets in Section 3.3. A discussion on the results is given in Section 3.4. Eventually, Section 3.6 is devoted to the explicit construction of kernels, which satisfy the assumptions we impose for the zero-crossing-time-technique.

3.1 The zero-crossing-time-technique

3.1.1 Model and assumptions

Suppose we have observations Y_i from the model

$$Y_i = m(x_i) + \varepsilon_i, \quad i \in \{1, \ldots, n\}, \tag{3.1}$$

where the covariates $x_i = x_{i,n}$ are deterministic, equidistant design points in $[0,1]$.

We impose the following assumptions.

Assumption 3.1 (Errors). The $\varepsilon_i = \varepsilon_{i,n}$ are centered, independent and identically distributed random variables with standard deviation $\sigma > 0$, and for any $u > 0$, $P(|\varepsilon_1| > u) \leq 2\exp(-3u^2/\sigma_g^2)$ for some $\sigma_g \geq \sigma$. ◊

Remark 4. Note that the factor 3 in the exponential function is only required for sake of convenience in order to use (without additional notational effort) the results in part A.2 of the appendix, which are based on the work of Viens and Vizcarra (2007).

Given $\theta \in (0,1)$ and a continuous function $g : [0,1] \setminus \{\theta\} \to \mathbb{R}$ defined on $[0,1]$ except at θ, we denote the one-sided limits of g at θ by $g(\theta+) = \lim_{x\downarrow\theta} g(x)$, $g(\theta-) = \lim_{x\uparrow\theta} g(x)$ if these limits exist. We let $[g](\theta) = g(\theta+) - g(\theta-)$ denote the jump-height at θ, assuming that the limits actually do exist. We write $C^k(\{\theta\}^c)$ for the k-times continuously differentiable functions on $[0,1] \setminus \{\theta\}$, and for $L > 0$ let

$$\mathrm{Lip}(\{\theta\}^c, L) = \big\{g \in C(\{\theta\}^c) \,|\, |g(x) - g(y)| \le L\,|x-y|,\ x,y \in [0,1] \setminus \{\theta\},\ x < y,\ \theta \notin (x,y)\big\}.$$

We shall assume that the regression function m in (3.1) is an element of the function class $\mathcal{M}_s$, defined as follows.

Definition 3.1 (Regression function). *Let $\gamma \in \mathbb{N}$, $s \in \mathbb{R}$ with $s \ge \gamma + 1$ and let $a, L > 0$ as well as $\Theta \subset (0,1)$ be a given compact set. Define the class of functions $m \in \mathcal{M}_s = \mathcal{M}_s(\gamma, a, \Theta, L)$ by assuming that $m \in C^{\gamma-1}[0,1]$, and that there is a unique $\theta_m \in \Theta$ such that*

(i) $m^{(\gamma-1)} \in C^1(\{\theta_m\}^c)$, and the jump height $[m^{(\gamma)}] := [m^{(\gamma)}](\theta_m)$ of $m^{(\gamma)}$ at θ_m satisfies $|[m^{(\gamma)}]| \ge a$,

(iia) in case $s = \gamma + 1$, we have that $m^{(\gamma)} \in \mathrm{Lip}(\{\theta_m\}^c, L)$,

(iib) in case $s > \gamma + 1$ we actually assume that $m^{(\gamma-1)} \in C^2(\{\theta_m\}^c)$ with $[m^{(\gamma+1)}](\theta_m) = 0$, that is the jump height of $m^{(\gamma+1)}$ is zero at θ_m, and that for the function

$$g_m^{(\gamma)}(x) = \begin{cases} m^{(\gamma+1)}(x), & x \ne \theta_m, \\ m^{(\gamma+1)}(\theta_m+), & x = \theta_m \end{cases}$$

we have that $g_m^{(\gamma)} \in \mathcal{H}^{s-(\gamma+1)}([0,1], L)$, where $\mathcal{H}$ is defined as in (1.4).

Remark 5. This function class is an adaptation of the function classes as in Definition 1 and 2 in Goldenshluger et al. (2006), where Sobolev-smoothness of the smooth-extension $g_m^{(\gamma)}$ is replaced by Hölder-smoothness.

3.1.2 The estimators

In this section the estimators for kink-location and kink-magnitude are introduced. In particular, the characterization as a Z-estimate will be used for the kink-location estimate, instead of an M-estimate characterization as in Goldenshluger et al. (2006) or Cheng and Raimondo (2008).

Probe functional

The probe functional $\psi_{h,m}$ is defined in (1.10) for an appropriate kernel function $K : \mathbb{R} \to \mathbb{R}$, specified in Assumption 3.2 below. We use a Priestley-Chao-type estimator $\hat{\psi}_{h,n}$ as in (1.11) to estimate the probe functional.

Assumption 3.2 (Kernel). Given $\gamma \in \mathbb{N}$ and $s \ge \gamma + 1$, suppose that the kernel $K : \mathbb{R} \to \mathbb{R}$ has support $\mathrm{supp}(K) = [-1,1]$, is $(\gamma+5)$-times differentiable inside its support and satisfies the following properties:

(i) $K^{(j)}(-1) = K^{(j)}(1) = 0, \quad j = 0, \dots, \gamma+3$,

(ii) $K^{(1)}$ is an odd function, in particular $K^{(1)}(0) = 0$,

(iii) if $\lfloor s \rfloor \ge \gamma + 2$ then $\int_{-1}^{1} x^j K^{(1)}(x)\mathrm{d}x = 0$ for $j = 1, \dots, \lfloor s - \gamma - 2 \rfloor$,

(iv) there are $0 < q_* < q_l < 1$ such that $K^{(1)}(x) > 0$ for $x \in [-q_l, 0)$ and $K^{(1)}$ has a unique global maximum at $-q_*$,

(v) for some $x^* \in (0,1)$ and $c_2 > 0$ we have that $|K^{(1)}(x)| \geq c_2|x|$, $x \in [-x^*, x^*]$.

$\diamond$

Remark 6. These assumptions ensure an appropriate behavior of the probe functional near the kink-location. Assumption 3.2, (iv) in combination with (ii) guarantee a change of sign of the probe functional near the kink-location with a global maximum and a global minimum of the probe functional in a close neighborhood, while Assumption 3.2, (i), resp. Assumption 3.2, (iii) makes sure that the first derivative of K drives the behavior of the probe functional by integration by parts resp. negligibility of bias. Eventually, Assumption 3.2, (v) implies that the zero is unique and well-separated.
Moreover, these assumptions are stricter than those stated in Cheng and Raimondo (2008), see Section 3.6 for a detailed comparison as well as for an explicit construction of kernels satisfying Assumption 3.2 for $\gamma = 1$ and $s \geq \gamma + 1$.

Estimate of the location of a kink

The estimation of the kink-location consists of two stages. In the first stage an interval which contains the kink-location with high probability will be constructed, while in the second stage the kink-location is estimated by a zero of the empirical probe functional inside this interval. Assumption 3.2 will ensure a zero inside the interval of the first stage with high probability.

In the following we shall always impose Assumption 3.1 and assume that the regression function m in model (3.1) satisfies $m \in \mathcal{M}_s$ as specified in Definition 3.1, and that the fixed kernel K satisfies Assumption 3.2 with parameters q_*, q_l and x^* in (iv) and (v).

Given $h > 0$ we then let

$$t^* = t^*(h;m) = \theta_m + hq_*, \qquad t_* = t_*(h;m) = \theta_m - hq_*. \tag{3.2}$$

Lemma 3.2. *There is an $h_0 > 0$ with $\Theta \subset [h_0, 1-h_0]$ such that*

$$\text{for } h_0 \geq h > 0 \quad \text{there is a } \tilde{\theta} = \tilde{\theta}_{h,m} \in [t_*(h;m), t^*(h;m)] \quad \text{such that} \quad \psi_{h,m}(\tilde{\theta}) = 0. \tag{3.3}$$

Here h_0 can be chosen uniformly over $m \in \mathcal{M}_s$ and depending only on the kernel K as well as on the Lipschitz constant L and the set Θ of $\mathcal{M}_s$. In particular, we have that $|\tilde{\theta}_{h,m} - \theta_m| = O_{\mathcal{M}_s}(h)$ for $h \in (0, h_0)$.

The proof is given in Section 3.5.1. Note that above and in the following we mean by dependence of some constant on the set Θ of $\mathcal{M}_s$ more precisely that the constant depends only on the maximum and minimum value of Θ.

Lemma 3.2 motivates the two stages of the estimation procedure: the interval $[t_*, t^*]$ contains the kink-location θ_m of m and a zero of the probe functional, while this zero is at least in an h neighborhood of θ_m. Thus, the unknown deterministic terms are replaced by empirical versions resp. appropriate estimates.

Turning to the definition of the estimate for an interval containing the kink-location θ_m of m, let

$$\begin{aligned} \hat{t}_* &= \hat{t}_*(h;n) := \min\{\arg\min_t \hat{\psi}_{h,n}(t), \arg\max_t \hat{\psi}_{h,n}(t)\}, \\ \hat{t}^* &= \hat{t}^*(h;n) := \max\{\arg\min_t \hat{\psi}_{h,n}(t), \arg\max_t \hat{\psi}_{h,n}(t)\}, \end{aligned} \tag{3.4}$$

and define the estimator for the kink-location by

$$\hat{\theta}_{h,n} \in \begin{cases} \{t \in [\hat{t}_*, \hat{t}^*] \mid \hat{\psi}_{h,n}(t) = 0\}, & \text{if the set is not empty,} \\ \{\frac{\hat{t}_*+\hat{t}^*}{2}\}, & \text{otherwise.} \end{cases} \tag{3.5}$$

From Lemma 3.14 the set $\{t \in [\hat{t}_*, \hat{t}^*] \mid \hat{\psi}_{h,n}(t) = 0\} \neq \emptyset$ with high probability and uniformly for $\mathcal{M}_s$, so that only this part of the definition is asymptotically relevant.

Estimate of the magnitude of a kink

For the height $[m^{(\gamma)}]$ in the kink, the expansion (3.27) for $i = 1$ (compare to (3.36)) suggests

$$\widetilde{[m^{(\gamma)}]}_h := h \frac{\psi^{(1)}_{h,m}(\tilde{\theta}_{h,m})}{(-1)^{\gamma+2} K^{(2)}(0)}. \tag{3.6}$$

Thus, an estimate for the kink-magnitude (we call it sometimes kink-size as well) is given by

$$\widehat{[m^{(\gamma)}]}_{h,n} := h \frac{\hat{\psi}^{(1)}_{h,n}(\hat{\theta}_{h,n})}{(-1)^{\gamma+2} K^{(2)}(0).}. \tag{3.7}$$

3.1.3 Rate of convergence

The following theorem is similar to Theorem 1 in Goldenshluger et al. (2006), however, we use a different proof technique and also derive the rate of convergence for the deterministic zero of the probe functional $\psi_{h,m}$. See the remark below for a discussion.

Theorem 3.3. *Consider model (3.1) and suppose Assumption 3.2 holds, then there exists an $h_0 > 0$ depending only on the kernel K, as well as on the Lipschitz constant L and the set Θ of $\mathcal{M}_s$ such that for any $h \in (0, h_0)$ holds*

$$|\tilde{\theta}_{h,m} - \theta_m| = O_{\mathcal{M}_s}(h^{s-\gamma+1}), \tag{3.8}$$

where $\tilde{\theta}_{h,m}$ as defined in (3.3). Further, suppose that Assumption 3.1 is also satisfied. There exist finite constants $h_0, C > 0$ depending only on the kernel K, the standard deviation of the noise σ and the sub-Gaussian parameter σ_g as in Assumption 3.1 as well as on the Lipschitz constant L and the set Θ of $\mathcal{M}_s$ such that if $h \in (0, h_0)$ and n are such that $nh^{2\gamma+1} \geq C\log(1/h)$ then

$$|\hat{\theta}_{h,n} - \theta_m| = O_{P,\mathcal{M}_s}(h^{s-\gamma+1}) + O_{P,\mathcal{M}_s}((nh^{2\gamma-1})^{-1/2}). \tag{3.9}$$

Moreover, the constants in the O-terms depend only on K, σ_g, L as well as s and are continuous in s.

The proof is provided in Section 3.5.4.

Remark 7.

1. The terms in (3.9) are of the same order if $h \cong n^{-1/(2s+1)}$ giving the rate $n^{-(s-\gamma+1)/(2s+1)}$ uniformly over $\mathcal{M}_s$.
2. The best possible rate of convergence in Theorem 3.3, i.e. using $h \cong n^{-1/2s+1}$, corresponds to the minimax rate obtained by Goldenshluger et al. (2006). To see this, note that $m-1$ (which is a smoothness parameter in their paper) resp. β in their setting corresponds to $s-(\gamma+1)$ resp.

γ in our setting. Thus, one has to substitute $m = s - \gamma$ and $\beta = \gamma$ to see that $n^{-(m+1)/(2m+2\beta+1)}$ in their setting is the same as $n^{-(s-\gamma+1)/(2s+1)}$ in our setting.

3. Following the lines of the proof of Theorem 1 in Goldenshluger et al. (2006) it is straightforward to verify the stronger result

$$\mathbb{E}_m|\theta_m - \hat{\theta}_{h,n}(h;f)|^2 = O_{\mathcal{M}_s}\left(h^{2(s-\gamma+1)}\right) + O_{\mathcal{M}_s}((nh^{2\gamma-1})^{-1}),$$

which certainly implies (3.9). However, as the estimate $\hat{\theta}_{h,n}$ is defined as a Z-estimate the convergence rate is a by-product of the asymptotic analysis.

Theorem 3.4. *Consider model (3.1) and suppose Assumption 3.2 holds, then there exists an $h_0 > 0$ depending only on the kernel K, as well as on the Lipschitz constant L and the set Θ of $\mathcal{M}_s$ such that for any $h \in (0, h_0)$ holds*

$$\left|\widetilde{[m^{(\gamma)}]}_h - [m^{(\gamma)}]\right| = O_{\mathcal{M}_s}(h^{s-\gamma}). \tag{3.10}$$

Further, suppose that Assumption 3.1 is satisfied. There exist finite constants $h_0, C > 0$ as in Theorem 3.3 such that if $h \in (0, h_0)$ and n are such that $nh^{2\gamma+1} \geq C\log(1/h)$ then

$$\left|\widehat{[m^{(\gamma)}]}_{h,n} - [m^{(\gamma)}]\right| = O_{P,\mathcal{M}_s}((nh^{2\gamma+1})^{-1/2}) + O_{P,\mathcal{M}_s}(h^{s-\gamma}). \tag{3.11}$$

Moreover, the constants in the O-terms depend only on K, σ_g, L as well as s and are continuous in s.

The proof is given in Section 3.5.4.

Remark 8. The terms in (3.11) are of the same order if $h \cong n^{-1/(2s+1)}$ for which we obtain the rate of convergence $n^{-(s-\gamma)/(2s+1)}$ uniformly over $\mathcal{M}_s$. This rate of convergence corresponds to the classic optimal nonparametric rate for estimation of $m^{(\gamma)}(x)$ when $m \in C^\gamma$, which gives reason to believe that this estimate is minimax optimal as well.

3.2 Asymptotic confidence sets

3.2.1 Asymptotic normality of the estimates

Suppose for this section that $\gamma \in \mathbb{N}$ and $s > 0$ are such that $s \geq \gamma + 1$.

The next theorem shows joint asymptotic normality of the kink-location estimate in (3.5) and the kink-size estimate in (3.7) around their deterministic counterparts (3.3) resp. (3.6).

Theorem 3.5. *In model (3.1) under the Assumptions 3.1 and 3.2 as well as if h and n are such that*

$$nh^{2\gamma+1}\log(1/h)^{-1} \to \infty, \quad \text{and} \quad nh^{4s-2\gamma+1} \to 0.$$

Then, for any $\mathbf{x} \in \mathbb{R}^2$

$$\sup_{m\in\mathcal{M}_s}\left|P_m\left[\begin{pmatrix} \tilde{w}_n^{loc}(h)^{-1}(\hat{\theta}_{h,n} - \tilde{\theta}_{h,m}) \\ \tilde{w}_n^{mag}(h)^{-1}\left(\widehat{[m^{(\gamma)}]}_{h,n} - \widetilde{[m^{(\gamma)}]}_h\right)\end{pmatrix} \leq \mathbf{x}\right] - \Phi_2(\mathbf{x})\right| = o(1), \tag{3.12}$$

where Φ_2 denotes the cumulative distribution of the bivariate standard normal distribution and the asymptotic standard deviation for the kink-location estimate resp. kink-magnitude estimate is

$$\tilde{w}_n^{loc}(h) := \frac{\sigma||K^{(\gamma+2)}||_2}{\sqrt{nh^{2\gamma-1}}\,[m^{(\gamma)}]\,K^{(2)}(0)}, \quad \textit{resp.} \quad \tilde{w}_n^{mag}(h) := \frac{\sigma||K^{(\gamma+3)}||_2}{\sqrt{nh^{2\gamma+1}}\,K^{(2)}(0)}. \tag{3.13}$$

Section 3.5.5 is devoted to the proof.

Remark 9.

1. By a slight undersmoothing, that is choosing $h \cong n^{-1/(2s+1)}\log(n)^{\zeta}$ for $\zeta < 0$ in view of Theorem 3.3 and Theorem 3.4, the asymptotic normality in (3.12) even holds uniformly over $\mathcal{M}_s$ for $\tilde{\theta}_{h,m}$ replaced by θ_m and $\widetilde{[m^{(\gamma)}]}_h$ by $[m^{(\gamma)}]$.
2. Theorem 3.5 reveals that the estimates are asymptotically independent, a surprising fact concerning that $\widehat{[m^{(\gamma)}]}_{h,n}$ is defined through $\hat{\theta}_{h,n}$.

Note that the results of Theorem 3.5 can already be used to construct honest confidence sets for the actual parameters $(\theta_m, [m^{(\gamma)}])$ over $\mathcal{M}_s$ by employing an undersmoothing as in part one of Remark 9. However, in the next section we construct confidence sets which are even adaptive with respect to the smoothness parameter s of a slightly smaller subset of $\mathcal{M}_s$ for which the result in Theorem 3.5 will be crucial.

3.2.2 Adaptive confidence sets

In general, it is not possible to construct honest and at the same time adaptive confidence sets for a large function class such as $\mathcal{M} = \bigcup_{s\in\mathcal{S}} \mathcal{M}_s(\gamma, a, \Theta, L)$, where $\mathcal{S} \subset \mathbb{R}_+$ is compact, see Section 1.4 for a discussion on this topic. However, Giné and Nickl (2010) provided a technique to construct adaptive confidence bands by slightly reducing the function class over which the union is build, which inspired our following approach.

Let $\gamma \in \mathbb{N}$ and $\underline{s}, \overline{s} \in \mathbb{R}_+$ be such that $\gamma + 1 \leq \underline{s} < \overline{s}$. Further, we choose integers $k_{\min,n}$ and $k_{\max,n}$ such that

$$2^{-k_{\min,n}} \cong \left(\frac{\log(n)}{n}\right)^{\frac{1}{2\overline{s}+1}}, \quad 2^{-k_{\max,n}} \cong \left(\frac{\log(n)^2}{n}\right)^{\frac{1}{2\gamma+1}}, \tag{3.14}$$

and set $\mathcal{K}_n = [k_{\min,n}, k_{\max,n}] \cap \mathbb{N}$ as well as

$$h_k = 2^{-k}, \quad k \in \mathcal{K}_n. \tag{3.15}$$

Definition 3.6. *Let $\gamma, k_0 \in \mathbb{N}$ and $\underline{s}, \overline{s}, b_1, b_2 \in \mathbb{R}_+$ be such that $\underline{s} < \overline{s}$ and $\underline{s} \geq \gamma + 1$ as well as $b_1 < b_2$. Further, let $a, L > 0$ and $\Theta \subset (0,1)$ is a compact set. Then, define*

$$\widetilde{\mathcal{M}} := \widetilde{\mathcal{M}}(\underline{s}, \overline{s}, b_1, b_2, k_0, \gamma, a, \Theta, L) = \bigcup_{s\in[\underline{s},\overline{s}]} \widetilde{\mathcal{M}_s}(b_1, b_2, k_0, \gamma, a, \Theta, L), \tag{3.16}$$

where

$$\begin{aligned} \widetilde{\mathcal{M}_s} &= \widetilde{\mathcal{M}_s}(b_1, b_2, k_0, \gamma, a, \Theta, L) \\ &:= \{m \in \mathcal{M}_s(\gamma, a, \Theta, L) \,|\, b_1 h_k^{s-\gamma} \leq |\psi_{h_k,m}(\theta_m)| \leq b_2 h_k^{s-\gamma}\ \forall k \geq k_0\} \end{aligned} \tag{3.17}$$

and the kernel for the probe functional inside (3.17) satisfies Assumption 3.2 with γ and $\overline{s}+1$. For any $m \in \widetilde{\mathcal{M}}$ let s_m be the unique smoothness parameter s for which m fulfills the bias condition in (3.17).

Subsamples

We divide the sample into two subsamples. The distinction will be used for the adaptivity of the constructed confidence intervals, as one subsample serves for the estimation of the kink-location and the kink-size, while the other subsample acts for the estimation of the bandwidth via a Lepski-type choice. In particular, we make the following assumption on the subsamples.

Assumption 3.3 (Subsamples). The sample $S = \{Y_1,\dots,Y_n\}$ can be split into two parts, $S_1 = \{Y_1,Y_3\dots,Y_{n-1}\}$ and $S_2 = \{Y_2,Y_4,\dots,Y_n\}$, each of size $n_1 = n_2 = n/2$ such that in particular $n_j \cong n$ for $j = 1,2$. $\diamond$

Due to the design in model (3.1) each subsample contains every other data point in order to maintain the noised information about the regression function over the whole domain. This distinction of the sample differs from Giné and Nickl (2010) since we do not have an i.i.d. setting for density estimation.

Lepski-type choice of the resolution level

For a suitable constant $C_{Lep} > 0$ we choose the bandwidth based on the subsample S_2 as $h_{\hat{k}_n}$, where

$$\hat{k}_n = \min\{k \in \mathcal{K}_n \mid |\hat{\theta}_{h_k,n_2} - \hat{\theta}_{h_l,n_2}| \leq C_{Lep}\sqrt{\log(n_2)/n_2 h_l^{2\gamma-1}}, \forall\, l > k, \quad l \in \mathcal{K}_n\}, \tag{3.18}$$

which is a data-driven choice in the spirit of Lepski' method.

As it is typical for the construction of confidence intervals to employ an undersmoothing (see Section 1.4), we use a slightly higher resolved bandwidth than $h_{\hat{k}_n}$, that is $h_{\hat{k}_n+u_n}$ for the estimation of the kink-location θ_m and $h_{\hat{k}_n+v_n}$ for the estimation of the kink-size $[m^{(\gamma)}]$, where we have the following assumption on u_n resp. v_n (compare to Condition 4 in Giné and Nickl (2010)).

Assumption 3.4. Let u_n and v_n be integer-valued sequences such that $h_{u_n} \cong \log(n)^{-1/(2\gamma-1)}$ and $h_{v_n} \cong \log(n)^{-1/(2\gamma+1)}$. $\diamond$

The different assumptions on u_n and v_n are due to the distinct rates for estimating the kink-location resp. kink-size, see Theorems 3.3 and 3.4, such that different undersmoothing values are employed.

We turn to the construction of the adaptive confidence sets. Let $\hat{\sigma}$ be a uniform consistent estimate for σ. Define the following data-driven estimate for the asymptotic standard deviation of the kink-location estimate in (3.13)

$$\hat{w}_{n_1}^{loc}(h_{\hat{k}_n+u_n}) := \frac{\hat{\sigma}_{n_1}\,||K^{(\gamma+2)}||_2}{\sqrt{n_1 h_{\hat{k}_n+u_n}^{2\gamma-1}}\ \widehat{[m^{(\gamma)}]}_{h_{\hat{k}_n},n_1} K^{(2)}(0)}, \tag{3.19}$$

and similarly for the asymptotic standard deviation of the kink-magnitude estimate in (3.13) set

$$\hat{w}_{n_1}^{mag}(h_{\hat{k}_n+v_n}) := \frac{\hat{\sigma}_{n_1}\,||K^{(\gamma+3)}||_2}{\sqrt{n_1 h_{\hat{k}_n+v_n}^{2\gamma+1}}\ K^{(2)}(0)}. \tag{3.20}$$

For sake of brevity let us write $[x \pm a]$ for $[x-a, x+a]$ with $x \in \mathbb{R}$, $a > 0$. Then, for any $\alpha \in (0,1)$ we define the following confidence sets

$$\begin{aligned} C_n^{loc}(\alpha) &= [\hat{\theta}_{h_{\hat{k}_n+u_n},n_1} \pm \hat{w}_{n_1}^{loc}(h_{\hat{k}_n+u_n})\, q_{1-\alpha/2}(W)], \\ C_n^{mag}(\alpha) &= \left[\widehat{[m^{(\gamma)}]}_{h_{\hat{k}_n+v_n},n_1} \pm \hat{w}_{n_1}^{mag}(h_{\hat{k}_n+v_n})\, q_{1-\alpha/2}(W)\right], \end{aligned} \tag{3.21}$$

where $q_\beta(W)$ is the β-quantile of $W = \max\{X_1, X_2\}$ with X_1, X_2 are independent standard normal distributed random variables. The following theorem, which is proven in Section 3.5.7, shows adaptivity as well as asymptotic honesty of the confidence sets over $\widetilde{\mathcal{M}}$.

Theorem 3.7. *Consider model (3.1) and let Assumptions 3.1, 3.3 as well as 3.4 be satisfied and K is a kernel satisfying Assumption 3.2 with γ and $\overline{s}+1$. Furthermore, let $\hat{\sigma}_n$ be an estimate such that $|\hat{\sigma}_n - \sigma| = o_{p,\widetilde{\mathcal{M}}}(1)$. Then, for any $\alpha \in (0,1)$*

$$\liminf_{n} \inf_{m\in\widetilde{\mathcal{M}}} P_m\big((\theta_m,[m^{(\gamma)}])^T \in C_n^{loc}(\alpha)\times C_n^{mag}(\alpha)\big) \geq 1-\alpha. \tag{3.22}$$

Furthermore, there exists a finite constant $C>0$ which can be chosen uniformly in $\widetilde{\mathcal{M}}$ such that

$$\limsup_{n} \sup_{m\in\widetilde{\mathcal{M}}} \left[P_m\left(\hat{w}_{n_1}^{loc}(h_{\hat{k}_n+u_n}) \geq C\left(\frac{\log(n)}{n}\right)^{\frac{s_m-\gamma+1}{2s+1}}\right) + P_m\left(\hat{w}_{n_1}^{mag}(h_{\hat{k}_n+v_n}) \geq C\left(\frac{\log(n)}{n}\right)^{\frac{s_m-\gamma}{2s+1}}\right)\right] = 0, \tag{3.23}$$

where s_m is as in Definition 3.6.

Remark 10.

1. Equation (3.22) shows asymptotically honesty of the confidence sets, see (1.36), while the adaptivity of the confidence sets as defined in (1.35) is covered by (3.23).
2. At first glance it is surprising that although Theorem 3.5 shows asymptotically independence of the kink-location and kink-size estimate it is sufficient to choose the bandwidth resolution in (3.18) depending only on the kink-location estimate to obtain adaptive confidence intervals for the kink-size as well. This is due to the fact that both estimates balance their bias and variance for the same order of the bandwidth resolution, see Theorem 3.3 and Theorem 3.4.
3. The assumption on $\hat{\sigma}_n$ made in Theorem 3.7 is mild as for the proposed estimates in the literature such as in Hall et al. (1990) or Dette et al. (1998) it can be easily shown that they fulfill this assumption, if restricted to an interval away from the kink-location.
4. Define
$$\begin{aligned}\tilde{C}_n^{loc}(\alpha) &= [\hat{\theta}_{h_{\hat{k}_n+u_n},n_1} \pm \hat{w}_{n_1}^{loc}(h_{\hat{k}_n+u_n})q_{1-\alpha/2}(N(0,1))],\\ \tilde{C}_n^{mag}(\alpha) &= \left[\widehat{[m^{(\gamma)}]}_{h_{\hat{k}_n+v_n},n_1} \pm \hat{w}_{n_1}^{mag}(h_{\hat{k}_n+v_n})q_{1-\alpha/2}(N(0,1))\right],\end{aligned} \tag{3.24}$$
where $q_\beta(N(0,1))$ is the β quantile of a standard normal distribution. Then the proof of Theorem 3.7 can be easily modified to show that $\tilde{C}_n^{loc}$ is an adaptive confidence interval for θ_m over $\widetilde{\mathcal{M}}$ and the same is true for $\tilde{C}_n^{mag}$ with respect to $[m^{(\gamma)}]$.
5. It can be shown that the estimates by Lepski's method are adaptive over the larger function class
$$\mathcal{M} = \bigcup_{s\in[\underline{s},\overline{s}]} \mathcal{M}_s(\gamma,a,\Theta,L),$$
where $\mathcal{M}_s(\gamma,a,\Theta,L)$ as in Definition 3.1.

3.3 Finite sample simulations

In this section we investigate the finite sample properties of the proposed asymptotic confidence intervals for the kink-location $\tilde{C}_n^{loc}$ in (3.24). For this purpose the following analysis consists of three parts. The first scenario analyzes the properties for a finite sample of the suggested asymptotic confidence interval $\tilde{C}_n^{loc}$ in (3.24) as well as the estimate of the asymptotic standard deviation (3.19) for a kink of first order. In addition, we investigate the asymptotic bias in order to emphasize its negligibility. The second part compares the asymptotic confidence intervals for the kink-location

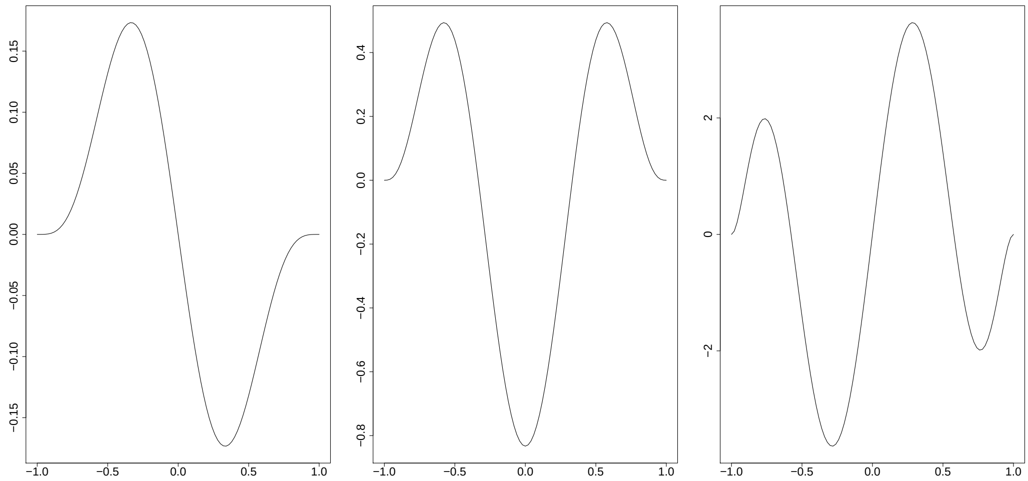

Figure 3.1.: From left to right: $K^{(1)}$, $K^{(2)}$ and $K^{(3)}$ given by (3.25).

$\tilde{C}_n^{loc}$ in (3.24) with the confidence intervals constructed by Mallik et al. (2013). Eventually, the third part of this section illustrates the applicability of our method to real-world datasets.

Concerning the kernel K, we have chosen it in all following scenarios such that

$$K^{(1)}(x) = \frac{8316}{832}\Big(-\frac{1}{12}x + \frac{1}{3}x^3 - \frac{1}{2}x^5 + \frac{1}{3}x^7 - \frac{1}{12}x^9\Big)1_{[-1,1]}(x). \tag{3.25}$$

In Figure 3.1 the first three derivatives of the used K are displayed. Note that only $K^{(i)}, i = 1,2,3$ are explicitly needed for the construction of the confidence sets for the location of a kink of first order.

First scenario

In this first scenario, we used two regression functions with different order of smoothness outside the kink. In particular, for the first regression function set $\theta_n = 1/2 + {}^1/_{3n}$ and define

$$m_1(x) = -2(x - \theta_n)1_{[0,\theta_n]}(x),$$

such that $[m_1^{(1)}] = -2$ and s is infinite in view of Definition 3.1 resp. Definition 3.6.

The second regression function m_2 is set to be the absolute value of the second anti-derivative of the Weierstraß-function (constrained on the unit interval). Let us write $m^{(-i)}$ for the i-th anti-derivative of a function m, then m_2 is given by

$$m_2(x) = |\tilde{m}_{9/10,7}^{(-2)}(x - {}^1/_{3n})|, \quad \tilde{m}_{c_1,c_2}(x) = \sum_{k=0}^{\infty} c_1^k \cos(c_2^k \pi x)1_{[0,1]}(x),$$

where $c_1 \in (0,1)$ and $c_2 \in 2\mathbb{N}+1$ such that $c_1c_2 > 1 + 3/2\pi$. It is a well known fact that the Weierstraß-function $\tilde{m}_{c_1,c_2}$ is everywhere continuous but not differentiable anywhere. Thus, m_2 is two times continuous differentiable except for the kink due to the absolute value, which is at $\theta_n = 1/2 - {}^1/_{3n}$ and therefore in this case $s = 2$ (in view of Definition 3.1 resp. Definition 3.6) and moreover $[m_2^{(1)}] \approx 9/4$.

In both scenarios we used $\varepsilon \sim SN(\zeta,\omega,\gamma)$, where $SN(\zeta,\omega,\gamma)$ denotes the skew-normal distribution with location parameter $\zeta \in \mathbb{R}$, scale-parameter $\omega > 0$ and shape parameter $\gamma \in \mathbb{R}$. For a given

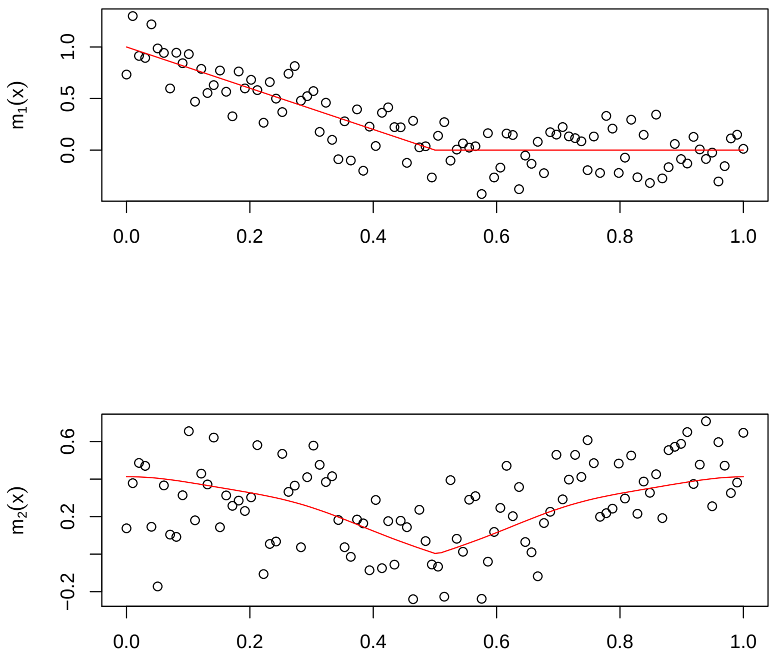

Figure 3.2.: Top: m_1 with noised observations. Bottom: m_2 with noised observations. The noise level is $\sigma = 0.2$ and the grid size $n = 100$ in both pictures.

noise-level $\sigma > 0$ (specified later), we chose the aforementioned parameters such that $\mathbb{E}\varepsilon = 0$ and $\mathbb{E}\varepsilon^2 = \sigma^2$. For illustration purposes consider Figure 3.2, where both regression functions and noised observations are displayed for a grid size of $n = 100$. The design was chosen in all settings such that the kink-location θ_n is not a design-point, respectively.

For the Lepski-scheme we have used a grid $\mathcal{K}_n$ for every scenario such that the bandwidth values are inside an interval $[h_{min,n}, h_{max,n}]$, where the values of $h_{min,n}$ resp. $h_{max,n}$ are given in Table 3.1. The Lepski-constant in (3.18) was chosen as $C_{Lep} = 0.09$ for m_1 and $C_{Lep} = 0.02$ for m_2.

Table 3.1.: Choice of $[h_{min,n}, h_{max,n}]$ for the first scenario.

	$n = 500$	$n = 1000$	$n = 2000$	$n = 4000$	$n = 8000$
m_1	[0.49,0.55]	[0.42,0.5]	[0.39,0.45]	[0.34,0.41]	[0.29,0.39]
m_2	[0.32,0.35]	[0.31,0.34]	[0.24,0.28]	[0.22,0.26]	[0.21,0.25]

All repeated simulations are based on a repetition number of 10000.

Asymptotic standard deviation estimation and kink-magnitude estimation

From (3.13) resp. (3.19) we see that the estimation of the asymptotic standard deviation is mainly driven by the estimation of the kink-magnitude and the estimation of the noise-level. A satisfying estimation of the noise-level can be achieved by using the simple noise-level estimator given in

Von Neumann (1941), that is

$$\hat{\sigma}_n = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} (Y_{i+1} - Y_i)^2.$$

As the regression functions are smooth outside the kink, one can even use more sophisticated estimates as in Hall et al. (1990) by restricting on intervals without a kink to obtain more precise results. However, the estimation of the noise-level is not a bottleneck for the finite sample simulation and works pretty well with the simple noise-level estimate above.

The results for the kink-magnitude estimation, i.e. the estimate $\widehat{[m^{(\gamma)}]}_{h_{\hat{k}_n+v_n}}$ in (3.7) are summarized in Table 3.2. As we can see, the square root of the Mean-Squared-Error (RMSE) is decreasing for increasing size of observations n for both scenarios, whereas for m_2 the estimation problem seems to be more difficult. For smaller samplesizes the estimator tends to overestimate the size of the kink, while this effect dwindles for larger samplesizes.

Table 3.2.: RMSE of the kink-magnitude estimate (3.7) based on the Lepski choice for $\sigma = 0.2$.

	$n = 500$	$n = 1000$	$n = 2000$	$n = 4000$	$n = 8000$
m_1	0.6227	0.1115	0.0721	0.0508	0.0348
m_2	1.7847	1.0466	0.7538	0.4109	0.2412

Moreover, in Table 3.3 the RMSE for the estimation of the asymptotic standard deviation for both scenarios are displayed as well as the asymptotic standard deviation for the respective setting. Again, the RMSE decreases and is of a quite small magnitude for both cases even though the estimation of the kink-magnitude for m_2 is not as accurate as for m_1.

Table 3.3.: RMSE for estimating the asymptotic standard deviation of the kink-location, i.e. $\tilde{w}_n^{loc}$ given in (3.13) for $\sigma = 0.2$.

	$n = 500$	$n = 1000$	$n = 2000$	$n = 4000$	$n = 8000$	$\tilde{w}_n^{loc}$
m_1	0.0094	0.0038	0.0027	0.0018	0.0012	0.3623
m_2	0.0899	0.0816	0.0525	0.0144	0.0062	0.3224

Performance of the confidence intervals

In this section we investigate the properties of the proposed confidence intervals for the kink-location in (3.24) with the noise-level- and kink-magnitude estimates as in the section above. The results are displayed in Table 3.4. It is evident that the empirical coverages are close to their nominal value and that the lengths of the confidence intervals become more slender with increasing size of observations n. An odd effect appears for m_2 by closer consideration of the case $n = 1000$ and $n = 2000$, that is the average width for $n = 2000$ is slightly larger than for $n = 1000$. This is due to the fact that on the one hand some very small estimates for the kink-size $[m_2^{(1)}]$ for $n = 2000$ appear, which results in some extremely large confidence intervals in some cases, and on the other hand using the empirical mean over the replications for the length computation. Using the empirical median instead, which is also a sensible choice for average length computation, this odd effect disappears. The top panels in Figure 3.3 resp. Figure 3.4 illustrate 100 outcomes of the constructed confidence intervals for the different sample sizes for the considered regression functions respectively. Apparently, the overall length shrinks up and the total number of outlying confidence intervals (confidence intervals which

Table 3.4.: Average coverage and length of the confidence intervals for the kink-location $\tilde{C}_n^{loc}$ in (3.21) for m_1 and m_2.

	90% nominal coverage		95% nominal coverage		99% nominal coverage	
	coverage	length	coverage	length	coverage	length
$n = 500$						
m_1	0.9030	0.0574	0.9490	0.0684	0.988	0.0899
m_2	0.8870	0.0703	0.9330	0.0838	0.978	0.1101
$n = 1000$						
m_1	0.8870	0.0530	0.9380	0.0631	0.9810	0.0829
m_2	0.9060	0.0511	0.9480	0.0609	0.9860	0.0801
$n = 2000$						
m_1	0.8870	0.0400	0.9410	0.0477	0.9860	0.0627
m_2	0.9000	0.0527	0.9430	0.0628	0.9830	0.0825
$n = 4000$						
m_1	0.8970	0.0298	0.9470	0.0355	0.9890	0.0466
m_2	0.9070	0.0337	0.9500	0.0402	0.9870	0.0528
$n = 8000$						
m_1	0.8960	0.0221	0.9470	0.0263	0.9900	0.0346
m_2	0.9010	0.0251	0.9550	0.0298	0.9880	0.0392

do not contain the true kink-location and are relatively far away from it) gets smaller too. Also the asymptotic normality of the Lepski estimate $\hat{\theta}_{h_{\hat{k}_n+u_n}}$ resp. $\widehat{[m^{(\gamma)}]}_{h_{\hat{k}_n+v_n}}$ is illustrated by the middle resp. bottom panel QQ-Plots of these Figures.

Negligibility of the asymptotic bias

As the confidence intervals are based on an undersmoothing, we consider the ratio between the computational bias and the standard deviation of the estimate $\hat{\theta}_{h_{\hat{k}_n+u_n}}$. Table 3.5 contains the results for our simulation. Apparently, the computed bias has a much smaller magnitude than the estimated standard deviation of the estimate. With increasing number of observations the ratio between both terms decreases in almost all cases, which suggests the negligibility of the bias even for finite sample considerations. In addition, it is not surprising that the ratios for m_2 are smaller than those for m_1, due to higher variability of the kink estimates, which is also visible by comparing the middle panels of Figure 3.3 and Figure 3.4. Worth noting is that the kink-locations in all setups were not one of the design points such that bias elimination by choice of the design was excluded.

Table 3.5.: Ratio between computed bias and computed standard deviation of the estimate $\hat{\theta}_{h_{\hat{k}_n+u_n}}$.

	$n = 500$	$n = 1000$	$n = 2000$	$n = 4000$	$n = 8000$
m_1	0.4337	0.0193	0.0026	0.0017	0.0110
m_2	0.0069	0.0036	0.0065	0.0035	0.0018

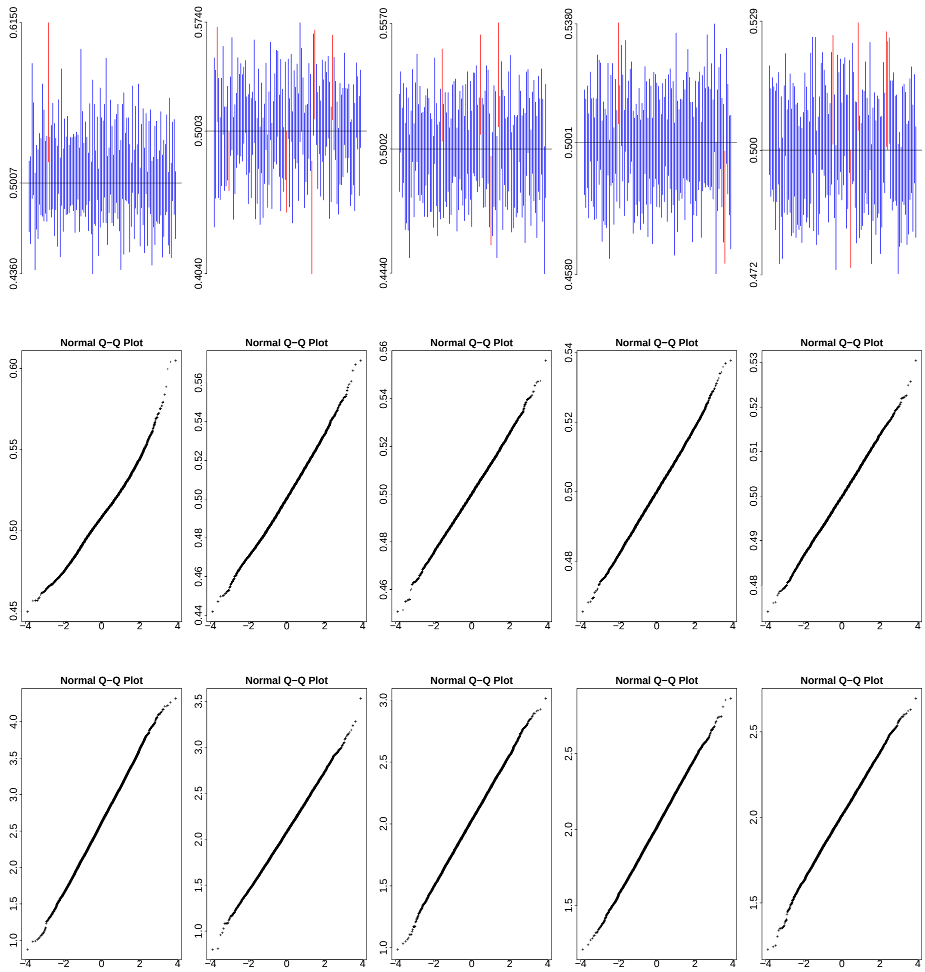

Figure 3.3.: Top panel: Illustration of the widths for 100 outcomes of the 95 % confidence intervals for θ_{m_1} for different sample sizes $n = 500, 1000, 2000, 4000, 8000$ respectively. Middle panel: QQ-Plots of the Lepski-estimator $\hat{\theta}_{h_{\hat{k}_n + u_n}}$ for m_1 for different sample sizes $n = 500, 1000, 2000, 4000, 8000$ respectively. Bottom panel: QQ-Plots of the Lepski-estimator for the size of the kink $\widehat{[m^{(\gamma)}]}_{h_{\hat{k}_n + v_n}}$ for m_1 for different sample sizes $n = 500, 1000, 2000, 4000, 8000$ respectively.

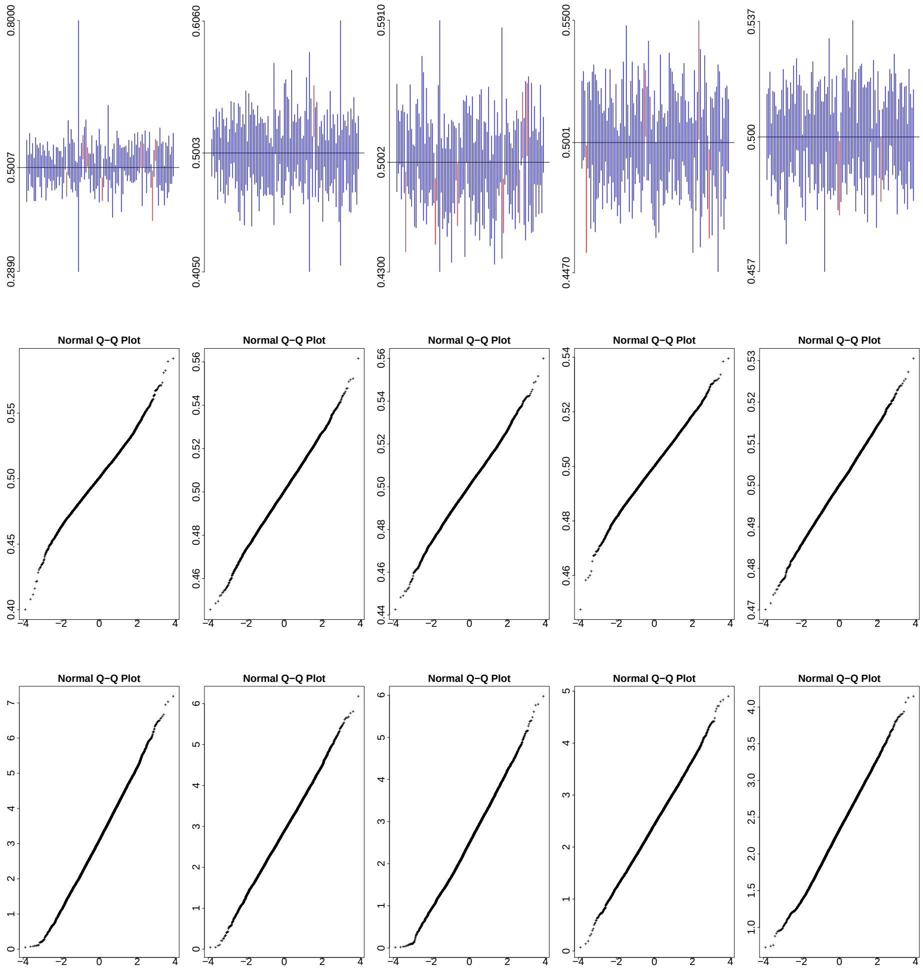

Figure 3.4.: Top panel: Illustration of the widths for 100 outcomes of the 95 % confidence intervals for θ_{m_2} for different sample sizes $n = 500, 1000, 2000, 4000, 8000$ respectively. Middle panel: QQ-Plots of the Lepski-estimator $\hat{\theta}_{h_{\hat{k}_n+u_n}}$ for m_2 for different sample sizes $n = 500, 1000, 2000, 4000, 8000$ respectively. Bottom panel: QQ-Plots of the Lepski-estimator for the magnitude of the kink $\widehat{[m^{(\gamma)}]}_{h_{\hat{k}_n+v_n}}$ for m_2 for different sample sizes $n = 500, 1000, 2000, 4000, 8000$ respectively.

Second scenario

We compared our proposed confidence intervals for the kink-location with those of Mallik et al. (2013) by simulating observations within the same setting as in Section 5 of their paper. In particular, we considered the regression functions $m_k(x) = (2(x-0.5))^k 1_{(0.5,1]}(x)$ for $k = 1,2$ and normally distributed noise variables with zero mean and standard deviation $\sigma = 0.1$. In Figure 3.5 the two regression functions and exemplary noised observations of them are displayed for the grid size $n = 100$. Both functions have a change-point in $\theta = 0.5$, whereas for m_1 this change-point corresponds

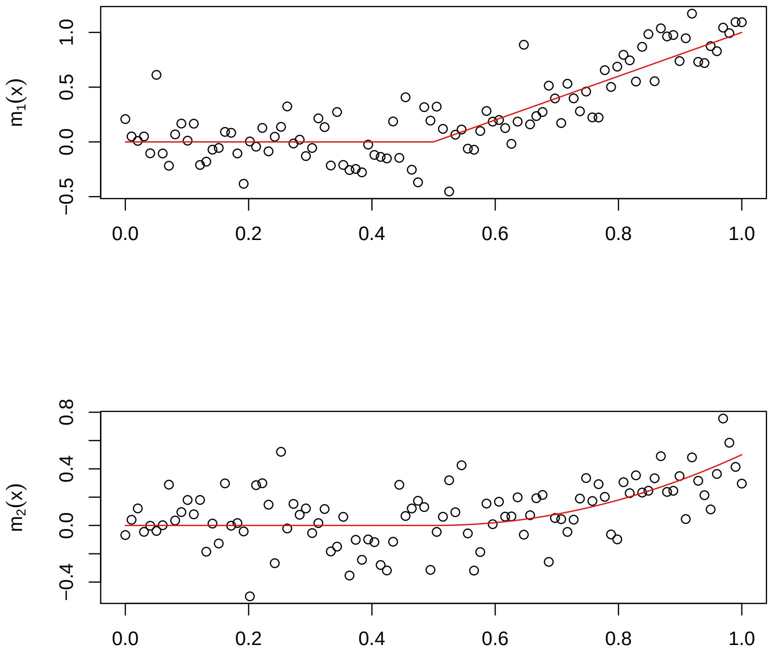

Figure 3.5.: Top: m_1 with noised observations. Bottom: m_2 with noised observations. The noise level is $\sigma = 0.1$ and the grid size $n = 100$ in both pictures.

to a kink of first order, while for m_2 the change-point is a kink of second order.
We applied our method in both settings over 5000 replications, as Mallik et al. (2013) did, where we used a grid $\mathcal{K}_n$ for every scenario such that the bandwidth values are inside an interval $[h_{min,n}, h_{max,n}]$, where the values of $h_{min,n}$ resp. $h_{max,n}$ are given in Table 3.6. For the Lepski-constant we used $C_{Lep} = 0.03$. The results can be found in Table 3.7 as well as the results of Mallik et al. (2013) for comparison. As expected, our method (denoted by OCI) yields confidence intervals which are

Table 3.6.: Choice of $[h_{min,n}, h_{max,n}]$ for the second scenario.

	$n = 100$	$n = 500$	$n = 1000$	$n = 2000$
m_1	[0.52,0.57]	[0.44,0.48]	[0.36,0.41]	[0.29,0.33]
m_2	[0.32,0.38]	[0.3,0.34]	[0.29,0.33]	[0.28,0.32]

narrower than those of Mallik et al. (2013) (denoted by MCI) since they have laxer assumptions on the smoothness of the regression function for their method. Especially, for m_1 the widths of our confidence sets are of a much smaller magnitude, while for m_2 the performance is slightly better.

Nevertheless, both m_1 and m_2 fulfill the assumptions of Mallik et al. (2013) as well as the assumptions for our setting and therefore, it seems reasonable to use our suggested confidence interval construction in such cases. Moreover, it is worth noting that the estimated bias for m_2 is of a small size and the confidence intervals become large due to the exponent $2\gamma-1$ for the bandwidth in (3.19), resulting in a relatively small denominator in the width term $\hat{w}_n^{loc}$.

Table 3.7.: Coverage probability and width (in parentheses) of the kink-location CI based on (3.21) (denoted by OCI) and the method of Mallik et al. (2013) (denoted by MCI) for the regression functions m_1 and m_2 and different grid sizes n.

m_1					
n	90 % CI		95 % CI		99 % CI
	MCI	OCI	MCI	OCI	OCI
100	0.939 (0.448)	0.811 (0.060)	0.972 (0.559)	0.883 (0.071)	0.962 (0.093)
500	0.922 (0.258)	0.888 (0.039)	0.965 (0.346)	0.943 (0.047)	0.987 (0.061)
1000	0.911 (0.197)	0.897 (0.030)	0.959 (0.265)	0.951 (0.036)	0.989 (0.047)
2000	0.903 (0.153)	0.896 (0.024)	0.954 (0.205)	0.946 (0.028)	0.985 (0.037)
m_2					
n	90 % CI		95 % CI		99 % CI
	MCI	OCI	MCI	OCI	OCI
100	0.883 (0.602)	0.909 (0.456)	0.917 (0.616)	0.948 (0.543)	0.987 (0.714)
500	0.889 (0.477)	0.891 (0.439)	0.934 (0.555)	0.941 (0.523)	0.983 (0.688)
1000	0.894 (0.424)	0.884 (0.415)	0.944 (0.525)	0.936 (0.495)	0.982 (0.650)
2000	0.899 (0.384)	0.905 (0.389)	0.948 (0.490)	0.945 (0.463)	0.983 (0.609)

Table 3.8.: Resulting confidence intervals $\tilde{C}_n^{loc}$ in (3.24) for different significance levels and corresponding kink-magnitude estimates and (rescaled) bandwidths for various choices of C_{Lep} for the global surface temperature dataset.

C_{Lep}	90%	95%	99%	$\widehat{[m^{(\gamma)}]}_{h_{\hat{k}_n+v_n}}$	$h_{\hat{k}_n}$
0.1	[1989,1990]	[1988,1990]	[1988,1991]	47.376	0.065
0.2 – 0.6	[1979,1989]	[1978,1990]	[1977,1991]	6.467	0.100
0.7	[1939,1944]	[1938,1945]	[1937,1946]	9.250	0.135
0.8	[1968,1976]	[1968,1977]	[1966,1978]	6.635	0.150
0.9	[1923,1935]	[1922,1936]	[1919,1938]	3.879	0.165
1.0	[1923,1935]	[1922,1936]	[1919,1938]	3.879	0.165
≥ 1.1	[1909,1914]	[1908,1914]	[1908,1915]	9.314	0.195

Third scenario: Real-world dataset applications

In this scenario we applied our proposed method to real-world datasets. For both datasets we used a grid $\mathcal{K}_n$ such that the range of bandwidths $[h_{min,n}, h_{max,n}]$ is [0.05,0.3]. Note that the datasets were scaled onto the unit interval for the application of our method.

Global surface temperature

In this application the dataset was the annual change in global surface temperature (in degree Celsius) for the time horizon 1880 to 2017 relative to the average temperatures for 1951 – 1980. This dataset is available at `https://data.giss.nasa.gov/gistemp`, where more details on the data are provided. For illustration purposes consider Figure 3.6. The investigation of such datasets is of special interest in today's world, as global warming is an issue attracting much public attention. For different choices of the Lepski-constant C_{Lep} in (3.18), we obtained the results summarized in Table 3.8, where we rounded the boundaries of the confidence intervals to obtain integer values. As

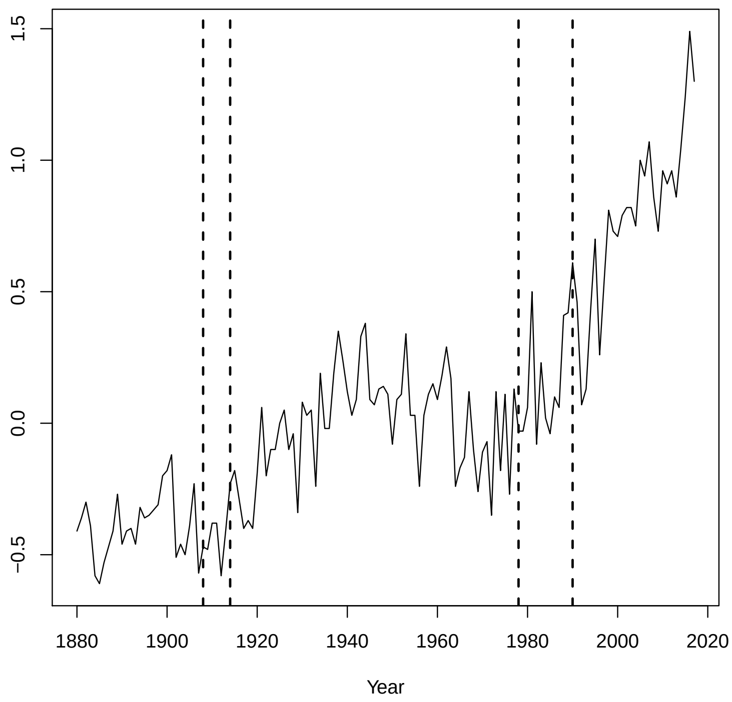

Figure 3.6.: Adaptive 95 % confidence intervals for the global surface temperature data for different choices of the Lepski constant in (3.18).

it can be seen, the confidence intervals vary with the choice of the Lepski constant. The smaller the Lepski constant was chosen, the more the center of the confidence intervals tended to the right boundary of the data. If the Lepski constant is chosen to be greater than 1.1 the detected kink is in accordance to Mallik et al. (2013), who detected with their method a kink at 1912 for a similar dataset. However, it seems reasonable to believe that another kink around 1984 appears in this data, since the plot suggests a steeper slope of the time series from that point on. For the detected kinks around 1912 and 1984 we displayed the corresponding confidence intervals based on our method in Figure 3.6.

Motorcycle data

For another application of our method to real-world data, we considered the motorcycle dataset, which was also investigated by Cheng and Raimondo (2008) for kink-detection purposes. The motorcycle data consists of 133 observations showing the effects of motorcycle crashes on victims heads, that is the acceleration of the head of a PTMO (post mortem human test object) depending on the time after a simulated motorcycle crash.
The data is displayed in Figure 3.7, while the results based on our method are summarized in Table 3.9. For this dataset only three possible kink estimates are detected for varying choices of the Lepski-constant C_{Lep}, though two of the estimates are in a similar location. Cheng and Raimondo (2008) detected kinks at similar positions and also pointed out that the dataset gives reason to believe that

a third kink around 15 time units is apparent. Presumably the other two detected kink-locations are

Table 3.9.: Resulting confidence intervals $\tilde{C}_n^{loc}$ in (3.24) for different significance levels and corresponding kink-magnitude estimates and (rescaled) bandwidths for various choices of C_{Lep} for the motorcycle dataset.

C_{Lep}	90%	95%	99%	$\widehat{[m^{(\gamma)}]}_{h_{\hat{k}_n+v_n}}$	$h_{\hat{k}_n}$
0.2	[30.615,32.241]	[30.459,32.397]	[30.155,32.701]	3303.155	0.060
[0.3,0.5]	[30.286,33.599]	[29.969,33.917]	[29.349,34.537]	1288.561	0.095
≥0.6	[20.996,21.688]	[20.930,21.755]	[20.800,21.884]	3499.251	0.295

of a such high magnitude, as indicated by the corresponding estimates in Table 3.9, that the method fails to detect a third kink-location.
In the right plot of Figure 3.7 the 95% confidence intervals $\tilde{C}_n^{loc}$ in (3.24) are illustrated by the shaded areas. This application shows that our method also works for non-equidistant design as well, since the time points are not uniformly spread in this dataset.

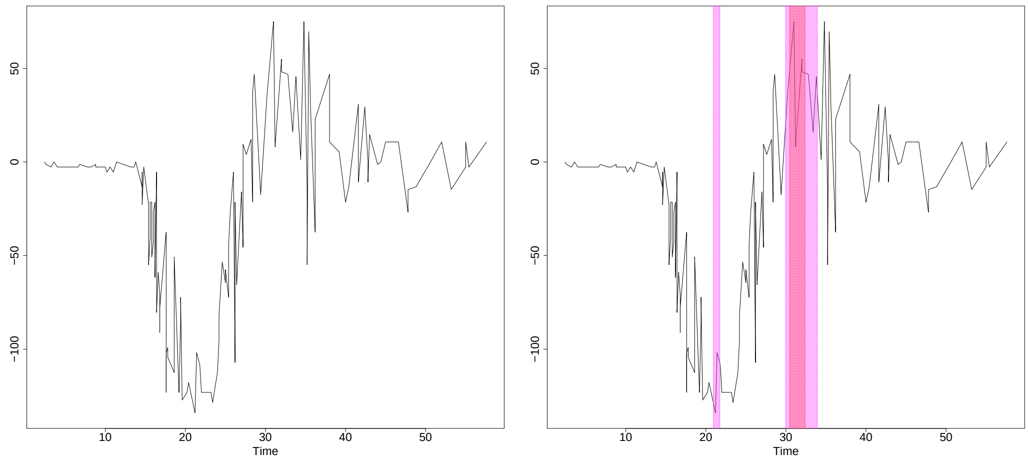

Figure 3.7.: Left: Illustration of the motorcycle data. Right: Shaded areas correspond to the adaptive 95 % confidence intervals for the three detected kink-locations based on different choices of the Lepski constant in (3.18).

3.4 Discussion

In this chapter we developed methods to construct adaptive asymptotic confidence sets for a single kink of higher order and its magnitude in an otherwise smooth regression function, for which to the best of our knowledge no methods were previously available. Several extensions are plausible due to our approach, which will be discussed in the following.

Bivariate extension

For the realm of image processing it would be of interest to extend the method to cope with images containing a single kink-location-curve in the spirit of the boundary fragment model considered in Chapter 2. However, it is not evident how to extend the zero-crossing-time-technique to a bivariate setting. A hybrid of the rotational difference kernel estimation and the zero-crossing-time-technique could serve for this purpose.

Several kinks

It is straightforward to extend the method to construct confidence sets for regression functions with several kinks, which have to be separated appropriately. Indeed, one can think of the model (3.1) with a regression function as in Definition 3.1 as a scaled fragment of the whole domain where a kink is expected. Decomposing the whole domain into such scaled fragments with suspected kinks and applying our suggested method with a Bonferroni correction for each of these scaled fragments would result in conservative confidence sets for the kink-locations.

Dependence considerations

Another appealing extension for the methods presented in this chapter is to consider different dependency structures in the data, such as in Wishart (2009) or Wishart and Kulik (2010). Although Wishart and Kulik (2010) analyzed the asymptotic distribution of their adapted criterion function, they did not analyze it of their kink-location estimate. Presumably, empirical process theory could serve for this purpose in their framework, as they represent their kink-location estimate $\hat{\theta}_n$ and the kink-location θ as $\hat{\theta}_n - \theta = Q_n(\lambda_n) - Q(\lambda)$, where Q is the cumulative quantile function of the design, while Q_n is its empirical version and λ_n resp. λ are surrogates for $\hat{\theta}_n$ resp. θ in a regular design setting.
For a long-range-dependency structure in the error sequence as in Wishart (2009) the challenging task would be to adapt the Lepski-scheme to deal with this dependency structure, whereas asymptotic normality should follow straightforwardly. The technical issue for the Lepski-method arising in a related context was already pointed out in Doukhan et al. (2002), see their discussion below Theorem 7.2.

Indirect setting

If the regression function m is only observed indirectly under convolution, (3.1) can be modeled as

$$Y_i = (\Upsilon * m)(x_i) + \varepsilon_i, \quad i \in \{1, \ldots, n\}, \tag{3.26}$$

where $(\Upsilon * m)(x) = \int\int \Upsilon(x-y)\, m(y)\,\mathrm{d}y$ for a point-spread function $\Upsilon : \mathbb{R} \to \mathbb{R}$ and x_i respectively ε_i as before. In this case the zero-crossing-time-technique in Section 3.1.2 has to be adapted in order to deconvolve the observations. Similar minimax lower bounds as in Goldenshluger et al. (2006) or Goldenshluger et al. (2008a) should hold for estimation of θ_m in (3.26) with additional assumptions on the kernel.

3.5 Proofs

If not otherwise indicated, we always consider model (3.1) and impose the Assumptions 3.1 and 3.2 for this section.

3.5.1 Properties of the probe functional

In the following we investigate the properties of the probe functional

$$\psi_{h,f}(t) = h^{-(\gamma+1)} \int_0^1 K^{(\gamma+2)}\big(h^{-1}(x-t)\big)\, m(x)\,\mathrm{d}x.$$

Lemma 3.8. *Let $m \in \mathcal{M}_s$ as given in Definition 3.1, then if $h_0 > 0$ is so small that $\Theta \subset [h_0, 1-h_0]$, then for $j = 0,1,2$ and for any $h \in (0, h_0)$ it holds that*

$$\begin{aligned} \psi_{h,m}^{(j)}(t) &= L_{h,j}(t) + O_{m \in \mathcal{M}_s, t \in [h,1-h]}(h^{s-\gamma-j}), \\ L_{h,j}(t) &= (-1)^{\gamma+1+j}\, h^{-j}\, [m^{(\gamma)}]\, K^{(1+j)}((\theta_m - t)/h). \end{aligned} \tag{3.27}$$

Moreover, the constant in the O-term depends only on the kernel K and on the Lipschitz constant L and the smoothness parameter s of $\mathcal{M}_s$ as in Definition 3.1, where the constant is continuous in s.

Proof of Lemma 3.8. Let h_0 be as in the assumption and $h \in (0, h_0)$. Given $t \in [h, 1-h]$ let $\tau = (\theta_m - t)/h$. By differentiation under the integral, substitution, γ-fold integration by parts (note that $m^{(\gamma)}$ is absolutely continuous) and Assumption 3.2, (i),

$$\begin{aligned} \psi_{h,m}^{(j)}(t) &= (-1)^j h^{-(\gamma+1+j)} \int_{[0,1]} m(x)\, K^{(\gamma+2+j)}(h^{-1}(x-t))\, \mathrm{d}x \\ &= (-1)^{\gamma+j}\, h^{-j} \int_{[-1,1]} m^{(\gamma)}(t+xh)\, K^{(2+j)}(x)\, \mathrm{d}x \\ &= (-1)^{\gamma+j}\, h^{-j} \int_{-1}^{\tau} m^{(\gamma)}(t+xh)\, K^{(2+j)}(x)\, \mathrm{d}x + (-1)^{\gamma+j} \int_{\tau}^{1} m^{(\gamma)}(t+xh)\, K^{(2+j)}(x)\, \mathrm{d}x. \end{aligned}$$

Further, since $K^{(1+j)}(-1) = K^{(1+j)}(1) = 0$ for $j = 0,1,2$ by Assumption 3.2, (i), we have that

$$\begin{aligned} (-1)^{\gamma+j}\, m^{(\gamma)}(\theta_m-) \int_{-1}^{\tau} K^{(2+j)}(x)\, \mathrm{d}x + (-1)^{\gamma+j}\, m^{(\gamma)}(\theta_m+) \int_{\tau}^{1} K^{(2+j)}(x)\, \mathrm{d}x& \\ = (-1)^{\gamma+j} K^{(1+j)}(\tau)\big[m^{(\gamma)}(\theta_m-) - m^{(\gamma)}(\theta_m+)\big]& \\ = (-1)^{\gamma+1+j}[m^{(\gamma)}]\, K^{(1+j)}(\tau).& \end{aligned}$$

Thus,

$$\begin{aligned} \psi_{h,m}^{(j)}(t) &= (-1)^{\gamma+1+j}\, h^{-j}[m^{(\gamma)}]\, K^{(1+j)}(\tau) + (-1)^{\gamma+j}\, h^{-j} \int_{-1}^{\tau} (m^{(\gamma)}(t+xh) - m^{(\gamma)}(\theta_m-))\, K^{(2+j)}(x)\, \mathrm{d}x \\ &\quad + (-1)^{\gamma+j}\, h^{-j} \int_{\tau}^{1} (m^{(\gamma)}(t+xh) - m^{(\gamma)}(\theta_m+))\, K^{(2+j)}(x)\, \mathrm{d}x \\ &=: (-1)^{\gamma+1+j}\, h^{-j}[m^{(\gamma)}]\, K^{(1+j)}(\tau) + J_{h,j}(t). \end{aligned}$$

Note that if $x \in (-1, \tau)$ then $t + xh < \theta_m$ and if $x \in (\tau, 1)$ then $t + xh > \theta_m$. Hence if $s - (\gamma+1) = 0$, from the Lipschitz continuity of $m^{(\gamma)}$ outside θ_m we directly obtain that $|J_{h,j}(t)| = O_{m \in \mathcal{M}_s, t \in [h,1-h]}(h^{s-\gamma-j})$. If $s - (\gamma+1) > 0$, then by integration by parts, $K^{(1+j)}(-1) = K^{(1+j)}(1) = 0$ for $j = 0,1,2$ and the definition of $g_m^{(\gamma)}$ we obtain

$$\begin{aligned} J_{h,j}(t) &= (-1)^{\gamma+1+j}\, h^{1-j} \int_{-1}^{1} g_m^{(\gamma)}(t+xh)\, K^{(1+j)}(x)\, \mathrm{d}x \\ &= (-1)^{\gamma+1+j}\, h^{1-j} \int_{-1}^{1} \big(g_m^{(\gamma)}(t+xh) - g_m^{(\gamma)}(t)\big)\, K^{(1+j)}(x)\, \mathrm{d}x. \end{aligned}$$

since $\int K^{(1+j)} = 0$ for $j = 0,1,2$. If $0 < s - (\gamma+1) \le 1$ we can directly use the uniform Lipschitz-continuity of $g_m^{(\gamma)}$ to get $|J_{h,j}(t)| = O_{m \in \mathcal{M}_s, t \in [h,1-h]}(h^{s-\gamma-j})$. While if $s - (\gamma+1) > 1$, obtain first by integration by parts, the vanishing moments and the vanishing edge properties of K in Assumption

3.2, (iii), that

$$\int_{-1}^{1} x^k K^{(1+j)}(x)\mathrm{d}x = 0, \quad \text{for} \quad k = 1,\ldots,\lfloor s-\gamma-2+j\rfloor. \tag{3.28}$$

Thus, by using Taylor expansion of $g_m^{(\gamma)}$ around t and (3.28), we also obtain that

$$|J_{h,j}(t)| = O_{m\in\mathcal{M}_s, t\in[h,1-h]}(h^{s-\gamma-j}).$$

Note that all the constants in the O-terms depend only on K, L as well as on s, where the constants are continuous in s, due to the remaining term in the Taylor expansion. Compare to classical results on bounding the bias term in nonparametric statistics, see for instance Proposition 1.2 in Tsybakov (2009). □

We show that for a kernel which satisfies Assumption 3.2 the probe functional has an approximate zero in θ_m which is also well-separated in a certain region. The following lemma is an adaptation of Lemma 2 in Goldenshluger et al. (2006), compare also to Lemma 1 in Cheng and Raimondo (2008).

Lemma 3.9 (Separation lemma)**.** *Let $h_0 > 0$ be so small that $\Theta \subset [h_0, 1-h_0]$ and $h \in (0, h_0.)$ Further, let $q \in (0, x^*)$, where x^* is as in Assumption 3.2, (v). Given $\delta \in (0, qh)$, let $A_{\delta,h,m} = \{t \mid \delta < |t-\theta_m| < qh\}$. Under Assumption 3.2, there are constants $C_i > 0$, $i = 1,2,3$, which can be chosen uniformly for $m \in \mathcal{M}_s$, and $t \in [h, 1-h]$ such that*

(i) $|\psi_{h,m}(\theta_m)| \le C_1 h^{s-\gamma}$,

(ii) if $\delta \ge C_2 h^{s-\gamma+1}$ *then*

$$\inf_{t\in A_{\delta,h,m}} (|\psi_{h,m}(t)| - |\psi_{h,m}(\theta_m)|) \ge C_3 \delta h^{-1}.$$

Moreover, the constants C_i depend only on the kernel K as well as on the Lipschitz constant L and the smoothness parameter s of $\mathcal{M}_s$ as in Definition 3.1, where the constants are continuous in s.

Proof of Lemma 3.9. *Ad(i).* Since $K^{(1)}(0) = 0$ we have $L_{h,0}(\theta_m) = 0$ in (3.27) and it follows that $|\psi_{h,m}(\theta_m)| \le C_1 h^{s-\gamma}$, where we choose C_1 as the constant for the $O_{m\in\mathcal{M}_s, t\in[h,1-h]}$-term in (3.27).

Ad (ii). Given $t \in A_{\delta,h,m}$, $\tau = (\theta_m - t)/h$ satisfies $q > |\tau| \ge \delta/h$. From Assumption 3.2, (v), it follows that $|K^{(1)}(\tau)|[m^{(\gamma)}] \ge c_2 |\tau| [m^{(\gamma)}] \ge c_2 h^{-1}\delta[m^{(\gamma)}]$, where c_2 is the kernel constant in Assumption 3.2, (v). From the assumption $\delta \ge C_3 h^{s-\gamma+1}$, (3.27) and the choice of C_1 it follows that

$$|\psi_{h,m}(t)| \ge c_2 h^{-1}\delta[m^{(\gamma)}] - C_1 h^{s-\gamma} \ge C_2 \delta h^{-1},$$

and in fact

$$\inf_{t\in A_{\delta,h,m}} (|\psi_{h,m}(t)| - |\psi_{h,m}(\theta_m)|) \ge C_2 \delta h^{-1}$$

for $C_2 := c_2 C_3 - 2C_1 > 0$ for sufficiently large C_3. Note that all the constants depend only on K, L as well as s and are continuous in s, due to Lemma 3.8. □

Proof of Lemma 3.2. We have for $L_{h,0}(t)$ in (3.27) that due to Assumption 3.2, (iv), $|L_{h,0}(t^*)| = |L_{h,0}(t_*)| = \big|[m^{(\gamma)}]\big|\big|K^{(1)}(-q_*)\big| > 0$, where $[m^{(\gamma)}] > 0$ is as in Assumption 3.1, and that $L_{h,0}(t^*)$ and $L_{h,0}(t_*)$ are of opposite signs since $K^{(1)}$ is odd by Assumption 3.2, (ii). Therefore, from (3.27) we have for sufficiently small $h_0 > 0$, depending only on K, L and Θ, that $\min\{|\psi_{h,m}(t^*)|, |\psi_{h,m}(t_*)|\} > 0$

and these are of opposite signs as well. The assertion follows from the continuity of the probe functional $\psi_{h,m}(t)$. □

From now on, we incorporate the empirical probe functional given by

$$\hat{\psi}_{n,h}(t) = n^{-1}h^{-(\gamma+1)}\sum_{i=1}^{n} Y_i K^{(\gamma+2)}(h^{-1}(x_i - t))$$

into our further analysis.

Lemma 3.10. *For the probe functional (1.10) and its empirical version (1.11), there exists a finite constant $h_0 > 0$ depending only on K and σ_g as in Assumption 3.1 as well as on L of $\mathcal{M}_s$, such that for any $h \in (0, h_0)$ and $n \in \mathbb{N}$ we have for $j = 0,1,2$ that*

(i) $\mathbb{E}_m\hat{\psi}^{(j)}_{h,n}(t) = \psi^{(j)}_{h,m}(t) + O_{m\in\mathcal{M}_s, t\in[0,1]}\big((nh^{\gamma+1+j})^{-1}\big)$, *for* $t \in [0,1]$,

(ii) $\sup_{t\in[0,1]}\big|\hat{\psi}^{(j)}_{h,n}(t) - \mathbb{E}_m\hat{\psi}^{(j)}_{h,n}(t)\big| = O_{P,\mathcal{M}_s}\Big(\Big(\frac{\log(1/h)}{nh^{2(\gamma+j)+1}}\Big)^{1/2}\Big)$.

Consequently, for $j = 0,1,2$ we have for any $h \in (0, h_0)$ and $n \in \mathbb{N}$ that

$$\sup_{t\in[0,1]} |\hat{\psi}^{(j)}_{h,n}(t) - \psi^{(j)}_{h,m}(t)| = O_{P,\mathcal{M}_s}\Big(\Big(\frac{\log(1/h)}{nh^{2(\gamma+j)+1}}\Big)^{1/2}\Big). \tag{3.29}$$

Moreover, the constants in the O-terms and h_0 depend only on K and σ_g as well as on the Lipschitz constant L of $\mathcal{M}_s$ as in Definition 3.1.

The discretization error contained in the remainder term in the lemma thus has the rate $(nh^{\gamma+1+j})^{-1}$ uniformly in $m \in \mathcal{M}_s$ and $t \in [0,1]$.

Proof of Lemma 3.10. *Ad (i).* It holds for $t \in [0,1]$ that

$$\begin{aligned}\mathbb{E}_m\hat{\psi}^{(j)}_{h,n}(t) &= n^{-1}h^{-(\gamma+1+j)}\sum_{i=1}^{n} m(x_i)K^{(\gamma+2+j)}(h^{-1}(x_i - t))\\ &= h^{-(\gamma+1+j)}\int m(x)K^{(\gamma+2+j)}(h^{-1}(x-t))\,\mathrm{d}x + R_n(t,h),\end{aligned}$$

where $R_n(t,h)$ is an error term of order $O_{m\in\mathcal{M}_s,t\in[0,1]}\big((nh^{\gamma+1+j})^{-1}\big)$, due to Riemann-sum approximation. Indeed, let $B(n,h)$ denote the index set for which the sum in the latter display is not zero. Due to the equidistant design in model (3.1) and the support of K in Assumption 3.2 it holds that $|B(n,h)| \le 2nh$. Let $A_i = [x_{i-1}, x_i)$ for $i = \{1,\dots,n-1\}$, where $x_0 = 0$ and $A_n = [x_{n-1}, x_n]$, then

$$\begin{aligned}R_n(t,h) &\le n^{-1}h^{-(\gamma+1+j)}\sum_{i\in B(n,h)} |\sup_{z_1\in A_i} m(z_1)K^{(\gamma+2+j)}(h^{-1}(z_1 - t)) - \inf_{z_2\in A_i} m(z_2)K^{(\gamma+2+j)}(h^{-1}(z_2 - t))|\\ &\le 4n^{-1}h^{-(\gamma+1+j)}C_{L,K},\end{aligned}$$

where $C_{L,K} > 0$ is the Lipschitz constant of the product of m and $K^{(\gamma+2+j)}$ which, by definition of $\mathcal{M}_s$, can be chosen uniformly in $m \in \mathcal{M}_s$ an depending only on K and L, which concludes *(i)*.

Ad (ii). Consider

$$\sqrt{nh^{2(\gamma+i)+1}}\big(\hat{\psi}^{(j)}_{h,n}(t) - \mathbb{E}_m\hat{\psi}^{(j)}_{h,n}(t)\big) = (nh)^{-1/2}\sum_{i=1}^{n}\varepsilon_i K^{(\gamma+2+j)}\big(h^{-1}(x_i - t)\big)$$

Then Lemma A.8 implies that there is a constant $C > 0$ depending only on K and σ_g such that

$$\mathbb{E}_m \sup_{t\in[0,1]} \left|\hat{\psi}^{(j)}_{h,n}(t) - \mathbb{E}_m \hat{\psi}^{(j)}_{h,n}(t)\right| \le C\sqrt{\log(1/h)}/\sqrt{nh^{2(\gamma+i)+1}},$$

provided $h_0 > 0$ is chosen appropriately (depending only on K and σ_g). The claim follows by Markov's inequality. Note that all the constants in the O-terms depend if necessary only on K, L and σ_g. □

In the following lemma we bound the variance of the empirical probe functional and its derivatives.

Lemma 3.11. *There exists an $h_0 > 0$ depending only on K such that for any $h \in (0, h_0)$, $t \in [h, 1-h]$ and $n \in \mathbb{N}$, we have that for $j = 0, 1, 2$,*

$$\mathsf{Var}_m(\hat{\psi}^{(j)}_{h,n}(t)) = n^{-1}h^{-2(\gamma+j)-1}\sigma^2 \| K^{(\gamma+2+j)} \|_2^2 + O_{m\in\mathcal{M}_s, t\in[0,1]}\big((nh^{\gamma+j+1})^{-2}\big).$$

Moreover, the constant in the O-term depends only on K and on the Lipschitz constant L of $\mathcal{M}_s$ as in Definition 3.1.

Proof. We compute

$$\begin{aligned}\mathsf{Var}_m(\hat{\psi}^{(j)}_{h,n}(t)) &= n^{-2}h^{-2(\gamma+j+1)}\sigma^2 \sum_{i=1}^n \left[K^{(\gamma+2+j)}(h^{-1}(x_i - t))\right]^2 \\ &= n^{-1}h^{-2(\gamma+j)-1}\sigma^2 \int \left[K^{(\gamma+2+j)}(x)\right]^2 \mathrm{d}x + O_{m\in\mathcal{M}_s, t\in[0,1]}\big((nh^{\gamma+j+1})^{-2}\big),\end{aligned}$$

where the order of the discretization error is derived as in the proof of Lemma 3.10, (i). □

Lemma 3.12. *There are finite constants $C, C_1, C_2, h_0 > 0$ which depend if necessary only on K, σ, σ_g and on the Lipschitz constant L of $\mathcal{M}_s$ such that if $h \in (0, h_0)$, $n \in \mathbb{N}$ and $\zeta_n > 0$ are such that*

$$\zeta_n\sqrt{nh^{2\gamma+1}}/\sigma - C/\sqrt{nh} \ge \zeta_n\sqrt{nh^{2\gamma+1}}/2\sigma > C_1\sqrt{\log(1/h)} > 0,$$

then

$$P_m\big(\sup_{t\in[0,1]} |\hat{\psi}_{h,n}(t) - \psi_{h,m}(t)| \ge \zeta_n\big) \le 2\exp\left(-C_2\zeta_n^2 nh^{2\gamma+1}\right), \quad m \in \mathcal{M}_s.$$

In particular, the constants h_0, C, C_1 and C_2 are uniformly for $m \in \mathcal{M}_s$.

Proof of Lemma 3.12. We let

$$Z_n(t;h) = \frac{1}{\sqrt{nh}\sigma}\sum_{i=1}^n \varepsilon_i K^{(\gamma+2)}(h^{-1}(x_i - t)), \qquad R_n(t;h) = \frac{1}{\sqrt{nh}\sigma}\big(\mathbb{E}_m\hat{\psi}_{h,n}(t) - \psi_{h,m}(t)\big).$$

Then

$$\frac{\sqrt{nh^{2\gamma+1}}}{\sigma}\big(\hat{\psi}_{h,n}(t) - \psi_{h,m}(t)\big) = Z_n(t;h) + R_n(t;h),$$

and choosing h_0 small enough (depending only on K and L) obtain by Lemma 3.10, (i), for any $t \in [0,1]$ that $|R_n(t;h)| \le C/\sqrt{nh}$, where $C > 0$ depends only on K and L. Thus, for an appropriate choice of the constants in the requirements of Lemma A.7 we have that

$$P_m\big(\sup_{t\in[0,1]}|\hat\psi_{h,n}(t)-\psi_{h,m}(t)|\geq\zeta_n\big)\leq P_m\Big(\sup_{t\in[0,1]}|Z_{n,1}(t)|\geq {}^{\zeta_n\sqrt{nh^{2\gamma+1}}}/\sigma - C/\sqrt{nh}\Big)$$
$$\leq 2\exp\left(-C_1\zeta_n^2 nh^{2\gamma+1}\right).$$

□

3.5.2 Uniform consistency of the first stage of the zero-crossing-time-technique

In this section we investigate the first stage of the zero-crossing-time-technique, that is the uniform consistency of the estimates in (3.4) against their deterministic counterparts in (3.2).

Lemma 3.13. *Let $q\in(0,x^*)$, then there are finite constants $h_0,C_1,C_2>0$ depending if necessary only on K,σ,σ_g as well as on the Lipschitz constant L, the set Θ and the smoothness parameter s of $\mathcal{M}_s$, such that if $h\in(0,h_0)$ and $n\in\mathbb{N}$ are such that $nh^{2\gamma+1}\geq C_1\log(1/h)$, then it holds that*

$$\max\left\{P_m(|\hat t_*(h;n)-t_*(h;m)|>{}^{hq}/2),P_m(|\hat t^*(h;n)-t^*(h;m)|>{}^{hq}/2)\right\}$$
$$\leq 2\exp\left(-C_2 nh^{2\gamma+1}\right).$$

Moreover, C_1,C_2 can be chosen continuously in s, while h_0 can be chosen independently of s. In particular,

$$|\hat t_*(h;n)-t_*(h;m)|=o_{P,\mathcal{M}_s}(h),\qquad |\hat t^*(h;n)-t^*(h;m)|=o_{P,\mathcal{M}_s}(h).$$

Proof of Lemma 3.13. We only show $P_m(|\hat t_*-t_*|>{}^{hq}/2)$, the other inequality can be derived analogously.

Case (i): Suppose that $(-1)^{\gamma+1}[m^{(\gamma)}]>0$.

Then by (3.27) for $j=0$ it holds for sufficiently small h_0 (depending on Θ) that $\psi_{h,m}(t_*)<0$ and in this case $\hat t_*=\arg\min_t\hat\psi_{h,n}(t)$. Hence, setting $B=\{t\in[0,1]\mid |t_*-t|>{}^{hq}/2\}$, we have that

$$\begin{aligned}P_m(|\hat t_*-t_*|>{}^{hq}/2)&\leq P_m\big(\exists t\in B:\ \hat\psi_{h,n}(t_*)\geq\hat\psi_{h,n}(t)\big)\\&=P_m\big(\exists t\in B:\ \hat\psi_{h,n}(t_*)-\psi_{h,m}(t_*)+\psi_{h,m}(t)-\hat\psi_{h,n}(t)\geq\psi_{h,m}(t)-\psi_{h,m}(t_*)\big)\\&\leq P_m\big(2\sup_{t\in[0,1]}|\psi_{h,m}(t)-\hat\psi_{h,n}(t)|\geq\inf_{t\in B}\left(\psi_{h,m}(t)-\psi_{h,m}(t_*)\right)\big).\end{aligned}$$

From Lemma 3.8, obtain for h_0 small enough (depending on Θ) that

$$\psi_{h,m}(t)-\psi_{h,m}(t_*)=(-1)^{\gamma+1}[m^{(\gamma)}]\left(K^{(1)}(h^{-1}(\theta_m-t))-K^{(1)}(h^{-1}(\theta_m-t_*))\right)+O_{m\in\mathcal{M}_s,t\in[h,1-h]}(h^{s-\gamma}),$$

so that for constants $\tilde C_i>0$ depending only on K,L and s we have that

$$\begin{aligned}\inf_{t\in B}\left(\psi_{h,m}(t)-\psi_{h,m}(t_*)\right)&\geq\tilde C_1\inf_{t\in B}(-1)^{\gamma+1}[m^{(\gamma)}]\left(K^{(1)}(h^{-1}(\theta_m-t))-K^{(1)}(h^{-1}(\theta_m-t_*))\right)-\tilde C_2\,h^{s-\gamma}\\&\geq\tilde C_1\inf_{x:|x-a_*|>q/2}\left(K^{(1)}(x)-K^{(1)}(a_*)\right)-\tilde C_2\,h^{s-\gamma}\geq\tilde C_3,\end{aligned}$$

where the second inequality follows by substitution and properties of $K^{(1)}$ and the last inequality is due to Lemma 3.9 by choosing h_0 appropriately (depending on Θ). Using Lemma 3.12, for sufficiently small h_0 (depending on K,σ,σ_g) and appropriate choice of C_1 in the assumption, there exists a constant $C_2 > 0$ such that

$$\begin{aligned} P_m(|\hat{t}_* - t_*| > {}^{hq}/2) &\le P_m(\sup_{t\in[0,1]} |\hat{\psi}_{h,n}(t) - \psi_{h,m}(t)| \ge \tilde{C}_3/2) \\ &\le 2\exp\left(-C_2 n h^{2\gamma+1}\right). \end{aligned}$$

Note that C_1 and C_2 can be chosen depending only on K,σ,σ_g,L as well as s and also continuous in s, due to Lemma 3.8 and 3.9, while the choice of h_0 is independent of s.

Case (ii): $(-1)^{\gamma+1}[m^{(\gamma)}] < 0$, then by (3.27) for $j = 0$ it holds for sufficiently small h_0 (depending on Θ) that $\psi_{h,m}(t_*) > 0$ and in this case $\hat{t}_* = \arg\max_t \hat{\psi}_{h,n}(t)$. Thus,

$$\begin{aligned} P_m(|\hat{t}_* - t_*| > {}^{hq}/2) &\le P_m\big(\exists t \in B: \hat{\psi}_{h,n}(t) \ge \hat{\psi}_{h,n}(t_*)\big) \\ &= P_m\big(\exists t \in B: \hat{\psi}_{h,n}(t) - \psi_{h,m}(t) + \psi_{h,m}(t_*) - \hat{\psi}_{h,n}(t_*) \ge \psi_{h,m}(t_*) - \psi_{h,m}(t)\big) \\ &\le P_m\big(2\sup_{t\in[0,1]} |\psi_{h,m}(t) - \hat{\psi}_{h,n}(t)| \ge \inf_{t\in B}\left(\psi_{h,m}(t_*) - \psi_{h,m}(t)\right)\big). \end{aligned}$$

We conclude with similar arguments as in case (i). □

Lemma 3.14. *There exist finite constants $h_0, C > 0$ depending if necessary only on K,σ,σ_g as well as on L,Θ and s of $\mathcal{M}_s$, such that if $h \in (0,h_0)$ and n are such that $nh^{2\gamma+1} \ge C\log(1/h)$, then*

1. *with high probability, uniformly in $m \in \mathcal{M}_s$ there exists a $\xi \in [\hat{t}_*,\hat{t}^*]$ such that $\hat{\psi}_{h,n}(\xi) = 0$,*
2. *we have that $|\hat{t}^* - \hat{t}_*| = O_{P,\mathcal{M}_s}(h)$,*
3. *with high probability, uniformly in $m \in \mathcal{M}_s$, we have that $\theta_m \in [\hat{t}_*,\hat{t}^*]$.*

Moreover, C can be chosen continuously in s, while h_0 is independent of s.

Proof of Lemma 3.14. Let $h_0 > 0$ be smaller or equal as all the h_0 terms in Lemma 3.2, 3.8, 3.12 and 3.13, which can be achieved by a choice depending only on K,σ,σ_g,L and Θ. Assume that $(-1)^{\gamma+1}[m^{(\gamma)}] > 0$ in which case $\psi_{h,m}(t_*) < 0$ and $\psi_{h,m}(t^*) > 0$. Let $\delta > 0$, then

$$\begin{aligned} P_m\big(\{\hat{\psi}_{h,n}(\hat{t}_*) \ge 0\}\big) \le{}& P_m\big(\{|\hat{\psi}_{h,n}(\hat{t}_*) - \psi_{h,m}(\hat{t}_*)| \ge \delta/2\}\big) + P_m\big(\{|\psi_{h,m}(\hat{t}_*) - \psi_{h,m}(t_*)| \ge \delta/2\}\big) \\ &+ 1_{\{\psi_{h,m}(t_*) \ge -\delta\}}. \end{aligned}$$

By choosing $\zeta_n = 1/\log(1/h)$ in Lemma 3.12, the first term tends to zero. By (3.27) it follows that $\psi_{h,m}$ is Lipschitz-continuous with constant of order h^{-1}. Hence the second term tends to zero by Lemma 3.13. Since we consider the case $\psi_{h,m}(t_*) < 0$, (compare to (3.27) for $j = 0$) the last term tends to zero for $\delta \to 0$. Hence $\hat{\psi}_{h,n}(\hat{t}_*) < 0$, and similarly $\hat{\psi}_{h,n}(\hat{t}^*) > 0$ with high probability, and the continuity of $\hat{\psi}_{h,n}$ implies statement 1., since all estimates hold uniformly over $m \in \mathcal{M}_s$.

Statement 2. follows since the distance between t_* and t^* is exactly of order h by definition (see (3.2)), and the distance $\hat{t}_* - t_*$ as well as $\hat{t}^* - t^*$ is of order $o_{P,\mathcal{M}_s}(h)$ by Lemma 3.13. Finally, since θ_m is at distance of order h both from t_* and t^*, statement 3. also follows from Lemma 3.13.

□

3.5.3 Exponential concentration inequality

The following lemma will be of great importance for the adaptive confidence sets. Note that the consistency respectively asymptotic normality of the estimates are shown without usage of the following lemma.

Lemma 3.15. *Let $\bar{C} > 0$ be some finite constant and $q \in (0, x^*)$, where x^* is as in Assumption 3.2, (v). There exist finite constants $C, C_1, C_2, h_0 > 0$ which depend if necessary only on K, σ, σ_g as well as on L, Θ and s of $\mathcal{M}_s$, such that if $h \in (0, h_0)$, $n \in \mathbb{N}$ and $\lambda_n > 0$ are such that*

$$\lambda_n \sqrt{nh^{2\gamma+1}}/\sigma - C/\sqrt{nh} \geq \lambda_n \sqrt{nh^{2\gamma+1}}/2\sigma > C_1 \sqrt{\log(1/h)} > 0,$$

and $\bar{C}\lambda_n/2 < qh$, then

$$\begin{aligned} &P_m(|\tilde{\theta}_{h,m} - \hat{\theta}_{h,n}| > \bar{C}\lambda_n) \\ &\qquad \leq 1_{\{|\tilde{\theta}_{h,m} - \theta_m| > \tau \bar{C}\lambda_n/2\}} + 2\exp(-C_2 \bar{C}^2 n h^{2\gamma-1} \lambda_n^2), \end{aligned}$$

where $\tau \in (0,1)$. Moreover, C, C_1, C_2 can be chosen continuously in s, while h_0 is independent of s.

Proof of Lemma 3.15. First, define the event

$$\Omega = \{|\hat{t}_*(h;n) - t_*| < {}^{hq}/_2\} \cap \{|\hat{t}^*(h;n) - t^*| < {}^{hq}/_2\}.$$

Then,

$$\begin{aligned} &P_m(|\tilde{\theta}_{h,m} - \hat{\theta}_{h,n}| > \bar{C}\lambda_n) \\ &\quad \leq 1_{\{|\tilde{\theta}_{h,m} - \theta_m| > \tau \bar{C}\lambda_n/2\}} + P_m(\{|\theta_m - \hat{\theta}_{h,n}| > {}^{(1-\tau)\bar{C}\lambda_n}/_2\} \cap \Omega) + P_m(\Omega^c). \end{aligned} \tag{3.30}$$

Lemma 3.13 implies for sufficiently small h_0 (depending only on K, σ, σ_g, L and on Θ)

$$P_m(\Omega^c) \leq 2\exp(-\tilde{C}_1 n h^{2\gamma+1}), \tag{3.31}$$

where $\tilde{C}_1 > 0$ is some finite constant uniform for $\mathcal{M}_s$ (depending only on K, σ, σ_g, L and on Θ). Let $\delta_n = {}^{(1-\tau)\bar{C}\lambda_n}/_2$, then on the event Ω holds that $|\theta_m - \hat{\theta}_{h,n}| < hq$, such that on Ω

$$\{|\theta_m - \hat{\theta}_{h,n}| > \delta_n\} \subset \{\exists t \in A_{\delta_n,h,m} : |\hat{\psi}_{n,h}(\theta_m)| \geq |\hat{\psi}_{n,h}(t)|\},$$

where $A_{\delta_n,h,m}$ is defined as in Lemma 3.9. The event $\{|\hat{\psi}_{n,h}(\theta_m)| \geq |\hat{\psi}_{n,h}(t)|\}$ can be rewritten to

$$\{|\hat{\psi}_{n,h}(\theta_m)| - |\psi_{h,m}(\theta_m)| + |\psi_{h,m}(t)| - |\hat{\psi}_{n,h}(t)| \geq |\psi_{h,m}(t)| - |\psi_{h,m}(\theta_m)|\}.$$

So it is easy to see that the latter event is contained in

$$\{2 \sup_{t\in[0,1]} |\hat{\psi}_{n,h}(t) - \psi_{h,m}(t)| \geq \inf_{t\in A_{\delta_n,h,m}} (|\psi_{h,m}(t)| - |\psi_{h,m}(\theta_m)|\}.$$

With Lemma 3.9, derive for appropriate choice of h_0 (depending only on Θ) that

$$\inf_{t\in A_{\delta_n,h,m}} (|\psi_{h,m}(t)| - |\psi_{h,m}(\theta_m)|\} \geq \tilde{C}_2 \lambda_n h^{-1}$$

for some constant $\tilde{C}_2 > 0$ which depends only on K, L as well as s and is continuous in s. Thus, by means of Lemma 3.12, for appropriate choice of the constants in the claim,

$$\begin{aligned} P_m(\{|\theta_m - \hat{\theta}_{h,n}| > \delta_n\} \cap \Omega) &\le P_m(2 \sup_{t\in[0,1]} |\hat{\psi}_{n,h}(t) - \psi_{h,f}(t)| \ge \tilde{C}_2\bar{C}\lambda_n h^{-1}) \\ &\le 2\exp(-\tilde{C}_3\bar{C}^2 n h^{2\gamma-1}\lambda_n^2) \end{aligned} \tag{3.32}$$

for some finite constant $\tilde{C}_3 > 0$ depending only on K, σ, σ_g, L as well as s and continuous in s. By assumption $\bar{C}\lambda_n/2 < qh$ so that one can find a suitable constant $C_2 > 0$ (depending only on K, σ, σ_g, L and on Θ) such that with (3.31) and (3.32)

$$\begin{aligned} &P_m(\Omega^c) + P_m(\{|\theta_m - \hat{\theta}_{h,n}| > \delta_n\} \cap \Omega) \\ &\qquad \le \exp(-C_1 n h^{2\gamma+1}) + \exp(-C_3\bar{C}^2 n h^{2\gamma-1}\lambda_n^2) \le \exp(-C_2\bar{C}^2 n h^{2\gamma-1}\lambda_n^2), \end{aligned}$$

which shows the first claim in view of (3.30). Eventually, note that the choice of h_0 did not dependent on s and furthermore C, C_1, C_2 were chosen depending only on K, σ, σ_g, L as well as s and also continuous in s, due to Lemma 3.8 and 3.9. □

3.5.4 Rates of convergence: outlines of the proofs of Theorems 3.3 and 3.4

In the following we shall restrict to the event that $\hat{\psi}_{h,n}(\hat{\theta}_{h,n}) = 0$ which is by Lemma 3.14 fulfilled with high probability and uniformly over $\mathcal{M}_s$ for appropriate choice of h_0. By Taylor expansion of $\psi_{h,m}$ at θ_m

$$0 = \psi_{h,m}(\tilde{\theta}_{h,m}) = \psi_{h,m}(\theta_m) + (\tilde{\theta}_{h,m} - \theta_m)\psi^{(1)}_{h,m}(\check{\theta}),$$

where $\check{\theta} = t\theta_m + (1-t)\tilde{\theta}_{h,m}$ for $t \in [0,1]$, is some value between θ_m and $\tilde{\theta}_{h,m}$. Similarly, by Taylor expansion of $\hat{\psi}_{h,n}$ at θ_m

$$0 = \hat{\psi}_{h,n}(\hat{\theta}_{h,n}) = \hat{\psi}_{h,n}(\theta_m) + (\hat{\theta}_{h,n} - \theta_m)\hat{\psi}^{(1)}_{h,n}(\ddot{\theta}),$$

where $\ddot{\theta} = \rho\theta_m + (1-\rho)\hat{\theta}_{h,n}$ for $\rho \in [0,1]$, is some (random) value between θ_m and $\hat{\theta}_{h,n}$. The preceding displays are equivalent to

$$\tilde{\theta}_{h,m} - \theta_m = -\frac{h\psi_{h,m}(\theta_m)}{h\psi^{(1)}_{h,m}(\check{\theta})}, \quad \hat{\theta}_{h,n} - \theta_m = -\frac{h\hat{\psi}_{h,n}(\theta_m)}{h\hat{\psi}^{(1)}_{h,n}(\ddot{\theta})}. \tag{3.33}$$

Asymptotics of the scale terms

The following lemma analyzes the asymptotic behavior of both denominator terms in (3.33).

Lemma 3.16. *Under the assumptions of Theorem 3.3, one has*

$$\begin{aligned} |h\psi^{(1)}_{h,m}(\check{\theta}) - (-1)^{\gamma+2}[m^{(\gamma)}]K^{(2)}(0)| &= o_{\mathcal{M}_s}(1), \\ |h\hat{\psi}^{(1)}_{h,n}(\ddot{\theta}) - (-1)^{\gamma+2}[m^{(\gamma)}]K^{(2)}(0)| &= o_{P,\mathcal{M}_s}(1). \end{aligned}$$

The proof is given in Section 3.5.6.

Convergence rates of the kink-location estimates: proof of Theorem 3.3

Proof of Theorem 3.3. Write $\hat{\psi}_{h,n}(\theta_m)$ as

$$\hat{\psi}_{h,n}(\theta_m) = \hat{\psi}_{h,n}(\theta_m) - \mathbb{E}\hat{\psi}_{h,n}(\theta_m) + \mathbb{E}\hat{\psi}_{h,n}(\theta_m).$$

Then, Lemma 3.10, (i) in combination with representation (3.27) for $j = 0$ in Lemma 3.8 yield for a suitable choice of h_0 (depending only on K,σ_g and L),

$$\mathbb{E}\hat{\psi}_{h,n}(\theta_m) = O_{\mathcal{M}_s}(h^{s-\gamma}) + O_{\mathcal{M}_s}\big((nh^{\gamma+1})^{-1}\big),$$

since $K^{(1)}(0) = 0$ due to Assumption 3.2, (ii). In addition, by suitable choice of h_0 (depending only on Θ), Lemma 3.11 and Chebychev's inequality imply $\hat{\psi}_{h,n}(\theta_m) - \mathbb{E}\hat{\psi}_{h,n}(\theta_m) = O_{P,\mathcal{M}_s}\big((nh^{2\gamma+1})^{-1/2}\big)$. Moreover, Lemma 3.16 implies that the denominator in the right term of (3.33) is uniformly over $\mathcal{M}_s$ a constant unequal zero, due to $K^{(2)} \neq 0$. Hence,

$$\frac{h\,\hat{\psi}_{h,n}(\theta_m)}{h\,\hat{\psi}^{(1)}_{h,n}(\ddot{\theta})} = O_{P,\mathcal{M}_s}(h^{s-\gamma+1}) + O_{P,\mathcal{M}_s}((nh^{2\gamma-1})^{-1/2}),$$

which yields the assertion for $\hat{\theta}_{h,n}$. A similar argumentation can be used to show (3.8) for the deterministic estimate $\tilde{\theta}_{h,m}$. Note that the constants in the O-terms depend only on K,σ_g,L as well as s and these constants can be chosen continuously in s, see Lemmas 3.8, 3.10 and 3.11. □

Convergence rates of the kink-size estimates: proof of Theorem 3.4

By Taylor expansion of $\psi^{(1)}_{h,m}$ in (3.6) at θ_m

$$\widetilde{[m^{(\gamma)}]}_h = h\,\frac{\big(\psi^{(1)}_{h,m}(\theta_m) + \psi^{(2)}_{h,m}(\bar{\theta})(\tilde{\theta}_{h,m} - \theta_m)\big)}{(-1)^{\gamma+2}K^{(2)}(0)}, \tag{3.34}$$

where $\bar{\theta}$ is between $\tilde{\theta}_{h,m}$ and θ_m. Likewise, by Taylor expansion of $\hat{\psi}^{(1)}_{h,n}$ in (3.7) around θ_m,

$$\widehat{[m^{(\gamma)}]}_{h,n} = h\,\frac{\big(\hat{\psi}^{(1)}_{h,n}(\theta_m) + \hat{\psi}^{(2)}_{h,n}(\grave{\theta})(\hat{\theta}_{h,n} - \theta_m)\big)}{(-1)^{\gamma+2}K^{(2)}(0)}, \tag{3.35}$$

where $\grave{\theta}$ is some value between θ_m and $\hat{\theta}_{h,n}$.

Proof of Theorem 3.4. For a suitable choice of h_0 (depending on Θ) it follows from (3.27) for $j = 1$ in Lemma 3.8 that

$$[m^{(\gamma)}] = (-1)^{\gamma+2}\,h\,\frac{\psi^{(1)}_{h,m}(\theta_m)}{K^{(2)}(0)} + O_{\mathcal{M}_s}(h^{s-\gamma}). \tag{3.36}$$

Subtracting this from (3.35) leads to

$$\big(\widehat{[m^{(\gamma)}]}_{h,n} - [m^{(\gamma)}]\big) = \frac{h\,\big(\hat{\psi}^{(1)}_{h,n}(\theta_m) - \psi^{(1)}_{h,m}(\theta_m)\big)}{(-1)^{\gamma+2}K^{(2)}(0)} + \frac{h\,\hat{\psi}^{(2)}_{h,n}(\grave{\theta})\,(\hat{\theta}_{h,n} - \theta_m)}{(-1)^{\gamma+2}K^{(2)}(0)} + O_{\mathcal{M}_s}(h^{s-\gamma}). \tag{3.37}$$

Now, Chebychev's inequality and Lemma 3.11 for $j = 1$ (for a suitable choice of h_0 depending on K) yield

$$h\,\big(\hat{\psi}^{(1)}_{h,n}(\theta_m) - \mathbb{E}\hat{\psi}^{(1)}_{h,n}(\theta_m)\big) = O_{P,\mathcal{M}_s}((nh^{2\gamma+1})^{-1/2}).$$

Further, by Lemma 3.10, (i), for $j = 1$ (for a suitable choice of h_0 depending if necessary on K and σ_g) derive

$$h\left(\mathbb{E}\hat{\psi}^{(1)}_{h,n}(\theta_m) - \psi^{(1)}_{h,m}(\theta_m)\right) = O_{\mathcal{M}_s}((nh^{\gamma+1})^{-1}),$$

such that the first term on the right-hand side in (3.37) is of order $O_{\mathcal{M}_s}((nh^{2\gamma+1})^{-1/2})$. Concerning the second term on the right-hand side of (3.37), Lemma 3.20 yields $h\hat{\psi}^{(2)}_{h,n}(\grave{\theta}) = o_{P,\mathcal{M}_s}(1)$, as $K^{(3)}(0) = 0$ by Assumption 3.2, (ii), as well as $|\grave{\theta} - \theta_m| = o_{P,\mathcal{M}_s}(h)$ by Lemma 3.19. This in combination with Theorem 3.3 implies that the second term on the right-hand side of (3.37) is $o_{P,\mathcal{M}_s}(1)$, which in summary leads to

$$\left|\widehat{[m^{(\gamma)}]}_{h,n} - [m^{(\gamma)}]\right| = O_{P,\mathcal{M}_s}((nh^{2\gamma+1})^{-1/2}) + O_{P,\mathcal{M}_s}(h^{s-\gamma}).$$

A similar argumentation shows the claim in (3.10) for the deterministic estimate $\widetilde{[m^{(\gamma)}]}_h$. Note that the constants in the O-terms depend only on K, σ_g, L as well as s and these constants can be chosen continuously in s, see Lemmas 3.8, 3.10 and 3.11. □

3.5.5 Asymptotic normality: outline of the proof of Theorem 3.5

Subtracting both terms in (3.33) and dividing by $\tilde{w}^{loc}_n$, as defined in (3.13), we get

$$\frac{\hat{\theta}_{h,n} - \tilde{\theta}_{h,m}}{\tilde{w}^{loc}_n(h)} = \frac{[m^{(\gamma)}]\,K^{(2)}(0)}{h\hat{\psi}^{(1)}_{h,n}(\ddot{\theta})}\hat{S}_1(h,n) + R_1(h,n), \tag{3.38}$$

where

$$\hat{S}_1(h,n) := \frac{\sqrt{nh^{2\gamma+1}}\,(\psi_{h,m}(\theta_m) - \hat{\psi}_{h,n}(\theta_m))}{\sigma\,||\,K^{(\gamma+2)}\,||_2} \tag{3.39}$$

and

$$R_1(h,n) := \frac{h\,[\hat{\psi}^{(1)}_{h,n}(\ddot{\theta}) - \psi^{(1)}_{h,m}(\breve{\theta})]\,\sqrt{nh^{2\gamma+1}}\,\psi_{h,m}(\theta_m)}{h^2\,\hat{\psi}^{(1)}_{h,n}(\ddot{\theta})\psi^{(1)}_{h,m}(\breve{\theta})} \times \frac{[m^{(\gamma)}]\,K^{(2)}(0)}{\sigma\,||\,K^{(\gamma+2)}\,||_2} \tag{3.40}$$

Likewise, subtracting (3.34) from (3.35) and dividing by $\tilde{w}^{mag}_n(h)$ leads to

$$\frac{\widehat{[m^{(\gamma)}]}_{h,n} - \widetilde{[m^{(\gamma)}]}_h}{\tilde{w}^{mag}_n(h)} = \hat{S}_2(h,n) + R_2(h,n). \tag{3.41}$$

where due to definition of $\tilde{w}^{mag}_n$ in (3.13) we have

$$\hat{S}_2(h,n) := \frac{\sqrt{nh^{2\gamma+3}}\,\left(\hat{\psi}^{(1)}_{h,n}(\theta_m) - \psi^{(1)}_{h,m}(\theta_m)\right)}{(-1)^{\gamma+2}\,\sigma\,||\,K^{\gamma+3}\,||_2} \tag{3.42}$$

and

$$R_2(h,n) := \frac{h^2\,\hat{\psi}^{(2)}_{h,n}(\grave{\theta})\,\sqrt{nh^{2\gamma-1}}\,(\hat{\theta}_{h,n} - \theta_m)}{(-1)^{\gamma+2}\,\sigma\,||\,K^{\gamma+3}\,||_2} - \frac{h^2\,\psi^{(2)}_{h,m}(\bar{\theta})\,\sqrt{nh^{2\gamma-1}}\,(\tilde{\theta}_{h,m} - \theta_m)}{(-1)^{\gamma+2}\,\sigma\,||\,K^{\gamma+3}\,||_2}. \tag{3.43}$$

The following lemma shows the negligibility of the remainder terms in (3.40) resp. (3.43).

Lemma 3.17. *Under the assumptions of Theorem 3.5, it holds*

$$\max\{|R_1(h,n)|, |R_2(h,n)|\} = o_{P,\mathcal{M}_s}(1).$$

The proof is given in Section 3.5.6.

The next lemma shows the joint asymptotic normality of the scores in (3.39) and (3.42).

Lemma 3.18. *Suppose the assumptions of Theorem 3.5 are fulfilled. Then, for any* $\mathbf{x} \in \mathbb{R}^2$

$$\sup_{m\in\mathcal{M}_s} \left| P_m\Big(\big(\hat{S}_1(h,n),\ \hat{S}_2(h,n)\big)^T \le \mathbf{x} \Big) - \Phi(\mathbf{x}) \right| = o(1).$$

The proof is provided in Section 3.5.6.

Proof of Theorem 3.5. Deduce with Lemma 3.16

$$\left| \frac{[m^{(\gamma)}]^2 K^{(2)}(0)^2}{h^2 \hat{\psi}^{(1)}_{h,n}(\ddot{\theta})^2} - 1 \right| = o_{P,\mathcal{M}_s}(1).$$

Hence, obtain by combination of Lemma 3.18 with the uniform Slutzky Theorem C.12 that for any $\mathbf{x} \in \mathbb{R}^2$,

$$\sup_{m\in\mathcal{M}_s} |P_m\Big(\big({}^{[m^{(\gamma)}]K^{(2)}(0)}/{}_{h\hat{\psi}^{(1)}_{h,n}(\ddot{\theta})}\, \hat{S}_1(h,n),\ \hat{S}_2(h,n)\big)^T \le \mathbf{x} \Big) - \Phi(\mathbf{x})| = o(1).$$

With this, Lemma 3.17 and the uniform Slutzky theorem C.12 we conclude the proof in view of (3.38) and (3.41). ◻

3.5.6 Proofs of auxiliary results in Section 3.5.4 and 3.5.5

Consistency of the kink-location estimate

By construction of t_* and t^* it holds that $|\theta_m - \tilde{\theta}_{h,m}| = O_{\mathcal{M}_s}(h)$, see Lemma 3.2. In addition, Lemma 3.14 implies $|\theta_m - \hat{\theta}_{h,n}| = O_{P,\mathcal{M}_s}(h)$. However, we need the following lemma to ensure a faster rate of convergence to analyze the term (3.33) for the proof of Theorem 3.3.

Lemma 3.19. *It holds that*

$$|\tilde{\theta}_{h,m} - \theta_m| = o_{\mathcal{M}_s}(h), \quad \textit{and} \quad |\hat{\theta}_{h,n} - \theta_m| = o_{P,\mathcal{M}_s}(h).$$

Proof of Lemma 3.19. Define the dilated criterion function $\bar{\psi}_{h,n}(w) := \hat{\psi}_{h,n}(\theta_m + hw)$ and for $m \in \mathcal{M}_s$ define $\hat{\Theta}_h = \{w \in \mathbb{R} \,|\, \theta_m + hw \in [\hat{t}_*, \hat{t}^*]\}$. Note that for the zeros $\hat{w}_{h,n}$ of $\bar{\psi}_{h,n}$ over $\hat{\Theta}_h$ and the zeros $\hat{\theta}_{h,n}$ of $\hat{\psi}_{h,n}$ it holds that $\hat{w}_{h,n} = (\hat{\theta}_{h,n} - \theta_m)/h$. For any $w \in [0,1]$ define

$$\psi(w;m) = (-1)^{\gamma+1} [m^{(\gamma)}] K^{(1)}(w).$$

Then, by (3.27) for $j = 0$ in Lemma 3.8

$$\sup_{w\in\hat{\Theta}_h} |\psi_{h,m}(\theta_m + hw) - \psi(w;m)| = o_{P,\mathcal{M}_s}(1),$$

since $\hat{\Theta}_h \subset \{w \in \mathbb{R} \,|\, \theta_m + hw \in [h, 1-h]\}$. Thus, Lemma 3.10 implies

$$\begin{aligned}\sup_{w\in\hat{\Theta}_h} |\bar{\psi}_{h,n}(w) - \psi(w;m)| &\leq \sup_{w\in\hat{\Theta}_h} |\bar{\psi}_{h,n}(w) - \psi_{h,m}(\theta_m + hw)| + \sup_{w\in\hat{\Theta}_h} |\psi_{h,m}(\theta_m + hw) - \psi(w;m)| \\ &= o_{P,\mathcal{M}_s}(1).\end{aligned}$$

Further, for any $\epsilon \in (0, x^*)$ Assumption 3.2, (v) yields

$$\inf_{m\in\mathcal{M}_s} \inf_{|w|>\epsilon} |\psi(w;m)| \geq \inf_{|w|>\epsilon} a c_2 |w| \geq a c_2 \epsilon > 0.$$

Apply Proposition 1.1, of which we have derived the assumptions in the latter two display by setting $\hat{g}_n = \bar{\psi}_{h,n}$ and $g_m = \psi$, to obtain

$$\frac{|\hat{\theta}_{h,n} - \theta_m|}{h} = |\hat{w}_{h,n}| = o_{P,\mathcal{M}_s}(1).$$

The assertion for $\tilde{\theta}_{h,m}$ follows analogously by noting that Proposition 1.1 is also true for non-random functions $\hat{g}_n$ and deterministic $\hat{w}_{h,n}$. □

Proof of Lemma 3.16

The following lemma immediately implies Lemma 3.16, due to Lemma 3.19.

Lemma 3.20. *Let $\check{\theta}, \ddot{\theta} \in \Theta$ and $\ddot{\theta}$ is random and $\check{\theta}$ is non-random. Then there exists an $h_0 > 0$ depending only on K, σ, σ_g, L and Θ such that if $h \in (0, h_0)$ and $n \in \mathbb{N}$, it holds for $j = 0, 1, 2$, that*

$$|h^j \psi_{h,m}^{(j)}(\check{\theta}) - (-1)^{\gamma+1+j}[m^{(\gamma)}] K^{(1+j)}(0)| = O_{\mathcal{M}_s}\left(h^{-1}|\check{\theta} - \theta_m|\right) + O_{\mathcal{M}_s}(h^{s-\gamma}),$$

$$|h^j \hat{\psi}_{h,n}^{(j)}(\ddot{\theta}) - (-1)^{\gamma+1+j}[m^{(\gamma)}] K^{(1+j)}(0)| = O_{P,\mathcal{M}_s}\left(h^{-1}|\ddot{\theta} - \theta_m|\right) + O_{P,\mathcal{M}_s}(h^{s-\gamma}) + O_{P,\mathcal{M}_s}\left(\sqrt{\frac{\log(n)}{nh^{2\gamma+1}}}\right).$$

Moreover, the constants in the O-terms depend only on the kernel K as well as on the Lipschitz constant L and the smoothness parameter s of $\mathcal{M}_s$ as in Definition 3.1, where the constants are continuous in s.

Proof of Lemma 3.20. Choosing h_0 appropriately, Lemma 3.8 implies

$$h^j \psi_{h,m}^{(j)}(\check{\theta}) = (-1)^{\gamma+1+j}[m^{(\gamma)}] K^{(1+j)}(h^{-1}(\theta_m - \check{\theta})) + O_{\mathcal{M}_s}(h^{s-\gamma}).$$

Now, by the mean value theorem

$$\begin{aligned}|K^{(1+j)}(h^{-1}(\theta_m - \check{\theta})) - K^{(1+j)}(0)| &= |K^{(1+j)}(h^{-1}(\theta_m - \check{\theta})) - K^{(1+j)}(h^{-1}(\theta_m - \theta_m))| \\ &\leq ||K^{(2+j)}||_\infty O_{P,\mathcal{M}_s}\left(h^{-1}|\check{\theta} - \theta_m|\right)\end{aligned}$$

which yields the first assertion. Similarly, by Lemma 3.8 and equation (3.29) for a suitable choice of h_0, obtain

$$\begin{aligned}h^j \hat{\psi}_{h,n}^{(j)}(\ddot{\theta}) &= h^j \psi_{h,m}^{(j)}(\ddot{\theta})) + O_{P,\mathcal{M}_s}\left(\sqrt{\frac{\log(n)}{nh^{2\gamma+1}}}\right) \\ &= (-1)^{\gamma+1+j}[m^{(\gamma)}] K^{(1+j)}(h^{-1}(\theta_m - \ddot{\theta})) + O_{\mathcal{M}_s}(h^{s-\gamma}) + O_{P,\mathcal{M}_s}\left(\sqrt{\frac{\log(n)}{nh^{2\gamma+1}}}\right).\end{aligned}$$

Now, using a similar argumentation as before with the mean value theorem it follows that

$$|K^{(1+j)}(h^{-1}(\theta_m - \ddot{\theta})) - K^{(1+j)}(0)| = O_{\mathcal{M}_s}(h^{-1}|\ddot{\theta} - \theta_m|),$$

which concludes the proof. Note that the constants in the O-terms depend only on K, σ_g, L as well as s and these constants can be chosen continuously in s, see Lemma 3.8. □

Neglibility of the remainder terms: proof of Lemma 3.17

Proof of Lemma 3.17. Let us start with $R_1(n,h)$ as defined in (3.40). Note that the second factor in (3.40) is a constant. Thus, we only need to investigate the first factor. By (3.29) for $j = 1$

$$h|\hat{\psi}^{(1)}_{h,n}(\ddot{\theta}) - \psi^{(1)}_{h,m}(\ddot{\theta})| = O_{P,\mathcal{M}_s}\Big(\sqrt{\frac{\log(n)}{nh^{2\gamma+1}}}\Big).$$

Next, recall Theorem 3.3 which implies

$$|\check{\theta} - \ddot{\theta}| = |t\theta_m + (1-t)\tilde{\theta}_{h,m} - (\rho\theta_m + (1-\rho)\hat{\theta}_{h,n})| = O_{P,\mathcal{M}_s}(h^{s-(\gamma-1)}) + O_{P,\mathcal{M}_s}((nh^{2\gamma-1})^{-1/2}),$$

where $\check{\theta}$ and $\ddot{\theta}$ as in (3.33). Therefore, by (3.27) for $j = 1$ deduce that

$$\begin{aligned} h|\psi^{(1)}_{h,m}(\ddot{\theta}) - \psi^{(1)}_{h,m}(\check{\theta})| &\le [m^{(\gamma)}]|K^{(2)}(h^{-1}(\theta_m - \ddot{\theta})) - K^{(2)}(h^{-1}(\theta_m - \check{\theta}))| + O_{\mathcal{M}_s}(h^{s-\gamma}) \\ &\le [m^{(\gamma)}]\|K^{(3)}\|_\infty \frac{|\check{\theta} - \ddot{\theta}|}{h} + O_{\mathcal{M}_s}(h^{s-\gamma}) \\ &= O_{P,\mathcal{M}_s}(h^{s-\gamma}) + O_{P,\mathcal{M}_s}((nh^{2\gamma+1})^{-1/2}). \end{aligned}$$

In summary, by the triangle inequality

$$h|\hat{\psi}^{(1)}_{h,n}(\ddot{\theta}) - \psi^{(1)}_{h,m}(\check{\theta})| = O_{P,\mathcal{M}_s}(h^{s-\gamma}) + O_{P,\mathcal{M}_s}\Big(\sqrt{\frac{\log(n)}{nh^{2\gamma+1}}}\Big).$$

Hence, the enumerator in $R_1(h,n)$ is of order

$$O_{P,\mathcal{M}_s}(\sqrt{n}h^{2s-\gamma+1/2}) + O_{P,\mathcal{M}_s}\Big(\sqrt{\log(n)h^{2(s-\gamma)}}\Big) = o_{P,\mathcal{M}_s}(1), \tag{3.44}$$

since $\sqrt{nh^{2\gamma+1}}\,\psi_{h,m}(\theta_m) = O_{\mathcal{M}_s}(\sqrt{n}h^{s+1/2})$ by Lemma 3.8 and the assumption on the asymptotics of h and n in Theorem 3.5. Finally, $R_1(h,n)$ is $o_{P,\mathcal{M}_s}(1)$ as the denominator is asymptotically a constant unequal to zero by Lemma 3.16 and Assumption 3.2, (ii).

Similarly, the terms in $R_2(h,n)$ are $o_{\mathcal{M}_s}(1)$ resp. $o_{P,\mathcal{M}_s}(1)$. To see this, we only analyze the enumerators in $R_2(h,n)$ as the denominators are both constant. Theorem 3.3 implies

$$\sqrt{nh^{2\gamma-1}}(\hat{\theta}_{h,n} - \theta_m) = O_{P,\mathcal{M}_s}(\sqrt{n}h^{s+1/2}) + O_{P,\mathcal{M}_s}(1).$$

Due to Lemma 3.20 and Theorem 3.3

$$h^2\hat{\psi}^{(2)}_{h,n}(\grave{\theta}) = O_{P,\mathcal{M}_s}(h^{s-\gamma}) + O_{P,\mathcal{M}_s}((nh^{2\gamma+1})^{-1/2})),$$

such that the first term in $R_2(h,n)$ is of the same order as in (3.44). A similar argumentation shows that the second term in $R_2(h,n)$ is $o_{\mathcal{M}_s}(1)$. □

Asymptotic normality of the score vector: proof of Lemma 3.18

Proof of Lemma 3.18. In this proof we also denote by $||\cdot||$ the euclidean norm on $\mathbb{R}^2$, which should lead not to confusion with the L_2-norm which is denoted in the same way. By Lemma 3.10, (i), for $j = 0, 1$, respectively, obtain

$$\begin{aligned}\sqrt{nh^{2(\gamma+j)+1}}\,(\psi^{(j)}_{h,m}(\theta_m) - \hat{\psi}^{(j)}_{h,n}(\theta_m)) &= (nh)^{-1/2}\sum_{i=1}^{n}\varepsilon_i K^{(\gamma+2+j)}\left(h^{-1}(x_i-\theta_m)\right) + O_{\mathcal{M}_s}((nh)^{-1/2}) \\ &=: E^{(j)}_n(m) + o_{\mathcal{M}_s}(1),\end{aligned}$$

due to the assumed asymptotics of h and n. By Theorem C.5 the term $\sqrt{nh^{2(\gamma+j)+1}}\,(\psi^{(j)}_{h,m}(\theta_m) - \hat{\psi}^{(j)}_{h,n}(\theta_m))$ and $E^{(j)}_n(m)$ have the same asymptotic limit distribution (provided it exists and satisfies the assumption of Theorem C.5) for $j = 0, 1$ respectively. For convenience set $\mathbf{E}_n(m) := \tilde{\Sigma}^{-1/2}\,(E^{(0)}_n(m), E^{(1)}_n(m))^T$, where

$$\tilde{\Sigma} = \sigma^2\begin{pmatrix} ||K^{\gamma+2}||_2^2 & 0 \\ 0 & ||K^{\gamma+3}||_2^2 \end{pmatrix}.$$

Hence, we show for any $\mathbf{x}\in\mathbb{R}^2$,

$$\sup_{m\in\mathcal{M}_s} |P_m(\mathbf{E}_n(m) \le \mathbf{x}) - \Phi_2(\mathbf{x})| = o(1), \tag{3.45}$$

which would conclude the proof. Note that $\mathbf{E}_n(m)$ depends on m only through θ_m, which is by definition of $\mathcal{M}_s$ element of Θ, a parameter of $\mathcal{M}_s$.

In order to prove (3.45), we intend to make use of the uniform version of the Lindeberg-Feller Theorem C.16, which can be applied since $\Phi_2(\cdot)$ does not depend on $\mathcal{M}_s$ and therefore $(\Phi_2(\cdot))_{m\in\mathcal{M}_s}$ fulfills the assumptions of the latter theorem. Therefore, we compute first the asymptotic covariance matrix of $\left(E^{(0)}_n(m), E^{(1)}_n(m)\right)^T$. By means of Lemma 3.11 for $j = 0, 1$, respectively, deduce

$$\begin{aligned}\mathsf{Var}_m(E^{(0)}_n(m)) &= \sigma^2||K^{(\gamma+2)}||_2^2 + o_{\mathcal{M}_s}(1),\\ \mathsf{Var}_m(E^{(1)}_n(m)) &= \sigma^2||K^{(\gamma+3)}||_2^2 + o_{\mathcal{M}_s}(1).\end{aligned}$$

Now, both $E^{(0)}_n(m)$ and $E^{(1)}_n(m)$ are centered such that their covariance is computed to be

$$\mathbb{E}_m\left(E^{(0)}_n(m)\,E^{(1)}_n(m)\right) = \frac{\sigma^2}{nh}\sum_{i=1}^{n}K^{(\gamma+2)}(h^{-1}(x_i-\theta_m))\,K^{(\gamma+3)}(h^{-1}(x_i-\theta_m)).$$

With a Riemann-sum approximation in a similar fashion as in the proof of Lemma 3.10 the latter term is

$$\begin{aligned}&\frac{\sigma^2}{h}\int K^{(\gamma+2)}(h^{-1}(x-\theta_m))\,K^{(\gamma+3)}(h^{-1}(x-\theta_m))\,\mathrm{d}x + O_{\mathcal{M}_s}((nh)^{-1})\\ &\qquad = \sigma^2\int K^{(\gamma+2)}(x)\,K^{(\gamma+3)}(x)\,\mathrm{d}x + o_{\mathcal{M}_s}(1) = o_{\mathcal{M}_s}(1),\end{aligned}$$

where the last equation holds since the function $x\mapsto K^{(\gamma+2)}(x)\,K^{(\gamma+3)}(x)$ is odd by Assumption 3.2, (ii). Thus,

$$\mathsf{Var}_m\left(\left(E^{(0)}_n(m), E^{(1)}_n(m)\right)^T\right) \to \tilde{\Sigma}$$

and the convergence holds uniformly over $\mathcal{M}_s$. Next, for any $\delta > 0$ we show that

$$\sup_{m\in\mathcal{M}_s}\sum_{i=1}^n ||\mathbf{a_i}(\theta_m)||_2^2\,\mathbb{E}_m\Big(\varepsilon_i^2 1_{\{||\mathbf{a_i}(\theta_m)||_2\,|\varepsilon_i|>\delta\}}\Big) = o(1),$$

where $\mathbf{a_i}(\theta_m) := (nh)^{-1/2}\big(K^{(\gamma+2)}(h^{-1}(x_i-\theta_m)), K^{(\gamma+3)}(h^{-1}(x_i-\theta_m))\big)^T$ and here $||\cdot||_2$ denotes the euclidean distance. Note that

$$\sup_{m\in\mathcal{M}_s}\max_{1\le i\le n} ||\mathbf{a_i}(\theta_m)||_2^2 \le (nh)^{-1}\max\{||K^{(\gamma+2)}||_\infty^2, ||K^{(\gamma+3)}||_\infty^2\} = o(1).$$

Further, the computation of the covariance matrix has shown that

$$\sum_{i=1}^n ||\mathbf{a_i}(\theta_m)||_2^2 \to ||K^{(\gamma+2)}||_2^2 + ||K^{(\gamma+3)}||_2^2 < \infty$$

and the convergence holds uniformly in $\mathcal{M}_s$. Hence, the former two displays lead us to

$$\begin{aligned}&\sup_{m\in\mathcal{M}_s}\sum_{i=1}^n ||\mathbf{a_i}(\theta_m)||_2^2\,\mathbb{E}_m\Big(\varepsilon_i^2 1_{\{||\mathbf{a_i}(\theta_m)||_2\,|\varepsilon_i|>\delta\}}\Big)\\ &\qquad\le \sup_{m\in\mathcal{M}_s}\max_{1\le i\le n}\mathbb{E}_m\Big(\varepsilon_i^2 1_{\{||\mathbf{a_i}(\theta_m)||_2\,|\varepsilon_i|>\delta\}}\Big)\sum_{i=1}^n ||\mathbf{a_i}(\theta_m)||_2^2 = o(1).\end{aligned}$$

□

3.5.7 Adaptive confidence sets: proof of Theorem 3.7

This section is devoted to the proof of Theorem 3.7, consequently suppose that the assumptions of Theorem 3.7 are satisfied.

Properties of the Lepski-type choice of the bandwidth

Define for b_1 and b_2, which are given in the definition of $\widetilde{\mathcal{M}}$ in (3.16), the following constants

$$\widetilde{b_1} := \frac{b_1}{[m^{(\gamma)}]K^{(2)}(0) - C_{b_1}}, \quad \text{and} \quad \widetilde{b_2} := \frac{b_2}{[m^{(\gamma)}]K^{(2)}(0) + C_{b_2}}, \tag{3.46}$$

where $a\,K^{(2)}(0) > C_{b_1}, C_{b_2} > 0$ are finite constants which are chosen below uniformly for $\widetilde{\mathcal{M}}$.

Recall the definitions of $\mathcal{K}_n$ resp. h_k in (3.14) resp. (3.15) and introduce for $s \in [\underline{s}, \overline{s}]$,

$$B(k,s) := \widetilde{b_2} h_k^{s-(\gamma-1)} = \widetilde{b_2}\, 2^{-k(s-\gamma+1)}, \quad \text{and} \quad \sigma(n,k) := \sqrt{\frac{k}{n\,h_k^{2\gamma-1}}} = \sqrt{\frac{2^{k(2\gamma-1)}\,k}{n}},$$

as well as

$$k_n^*(s) = \min\{k \in \mathcal{K}_n \,|\, B(k,s) \le C_{Lep}\sigma(n_2,k)/8\}.$$

The following lemma summarizes some properties of the just defined terms.

Lemma 3.21. *Let s_m be as defined in Definition 3.6.*

(i) $h_{k_{min,n}} > \ldots > h_{k_{max,n}}$.

(ii) $B(\cdot,s)$ *is decreasing, while* $\sigma(n_2,\cdot)$ *is increasing;*

(iii) $k_{min,n} \cong \log(n)$, $k_{max,n} \cong \log(n)$ *and* $k_{max,n} - k_{min,n} \cong \log(n)$.

(iv) $h_{k_n^*(s)} \cong \left(\log(n_2)/n_2\right)^{1/(2s+1)}$ *and in particular* $B(k_n^*(s),s) \cong \sigma\left(n_2,k_n^*(s)\right) \cong \left(\log(n_2)/n_2\right)^{(s-\gamma+1)/(2s+1)}$;

(v) $B(l,s) \leq C_{Lep}\sigma(n_2,k)/8$ *for* $k_n^*(s) \leq k \leq l \leq k_{max,n}$;

(vi) $h_k \geq h_{k_{max,n}} > \sigma(n_2,k)$ *for any* $k \in \mathcal{K}_n$;

(vii) $nh_k^{2\gamma+1}\log(1/h_k)^{-1} \longrightarrow \infty$ *for any* $k \in \mathcal{K}_n$;

(viii) for any $l \leq k \in \mathcal{K}_n$ *one has* $B(k,s) = 2^{(s-\gamma+1)(l-k)}B(l,s)$;

(ix) $\sigma(n_2,l) = \sigma(n_2,l+1)\left(2^{-\frac{2\gamma-1}{2}}\sqrt{l/(l+1)}\right)$.

Proof of Lemma 3.21. (i), (ii), (viii) and (ix) are easy. From (3.14) obtain

$$k_{\min,n} \cong \frac{\log(n/\log(n))}{\log(2)(2\overline{s}+1)}, \quad \text{and} \quad k_{\max,n} \cong \frac{\log(n/\log(n)^2)}{\log(2)(2\gamma+1)},$$

which immediately implies (iii). With this it is straightforward to obtain (iv) by balancing the terms in the definition of $k_n^*(s)$.
(v) follows by (iv) and (ii), as k and l are assumed to be greater or equal to $k_n^*(s)$. From (3.14) conclude that $\sigma(n_2,k_{\max,n}) \cong \sqrt{\log(n)}/n^{1/(2\gamma+1)}$, which is of a smaller order than $h_{k_{\max,n}}$. This shows (vi) due to (i) and (ii). Next, $nh^{2\gamma+1}_{k_{\max,n}}/\log(h^{-1}_{k_{\max,n}}) \cong \log(n)$, so that by (i) we conclude (vii). □

Part four of the lemma reveals that $k_n^*(s)$ is the index we would like to choose if $m \in \widetilde{\mathcal{M}_s}$. The next lemma shows that $\hat{k}_n$ defined in (3.18) is in some way a good estimate of $k_n^*(s)$ by appropriate choice of C_{Lep} in (3.18).

Lemma 3.22. *If* $C_{Lep} > 0$ *is chosen large enough depending only on* K,σ,σ_g *as well as on* L *and* Θ *of* $\widetilde{\mathcal{M}}$ *and if* n *is sufficiently large such that* $k_n^*(s_m) \geq 2$ *then there exists a* $\rho \in \mathbb{N}$ *depending only on* $b_1,b_2,\underline{s}$ *and on* C_{Lep} *such that it holds*

$$\sup_{m\in\widetilde{\mathcal{M}}} P_m(\hat{k}_n \notin [k_n^*(s_m)-\rho,k_n^*(s_m)]) = o(1),$$

where s_m *as in Definition 3.6.*

We need the following lemma for the proof of Lemma 3.22.

Lemma 3.23. *(i) If* C_{Lep} *is chosen large enough and depending only on* K,σ,σ_g *as well as on* L *and* Θ *of* $\widetilde{\mathcal{M}}$ *and if* n *is sufficiently large, then there exists a finite constant* $c_1 > 0$ *which is uniform in* $\widetilde{\mathcal{M}}$, *such that*

$$P_m(\hat{k}_n = k) \leq c_1 2^{-k/c_1}, \quad \forall k > k_n^*(s_m),$$

where s_m *as in Definition 3.6.*

(ii) If C_{Lep} is chosen large enough and depending only on K,σ,σ_g as well as on L and Θ of $\widetilde{\mathcal{M}}$, and also if n is large enough such that $k_n^*(s_m) \geq 2$, then there exist $\rho \in \mathbb{N}$ and $c_2 > 0$, which are both uniform in $\widetilde{\mathcal{M}}$, such that

$$P_m(\hat{k}_n = k) \leq c_2 2^{-k/c_2}, \quad \forall k < k_n^*(s_m) - \rho,$$

where s_m as in Definition 3.6. Moreover, ρ depends only on $b_1, b_2, \underline{s}$ and on C_{Lep}.

Proof of Lemma 3.22. Assume that C_{Lep} and n are sufficiently large, such that the statements of Lemma 3.23 hold. Let s_m be as in Definition 3.6. On the one hand, with Lemma 3.23, (i) it holds that

$$\sup_{m \in \widetilde{\mathcal{M}}} P_m(\hat{k}_n > k_n^*(s_m)) \leq \sup_{m \in \widetilde{\mathcal{M}}} \sum_{k_n^*(s_m) < k \leq k_{\max,n}} P_m(\hat{k}_n = k) \leq c_1 \sum_{k_n^*(s_m) < k \leq k_{\max,n}} 2^{-k/c_1} = o(1),$$

as c_1 is uniform in $\widetilde{\mathcal{M}}$ and due to the asymptotic behavior of the indices in $\mathcal{K}_n$, see Lemma 3.21, (iii). On the other hand, $P_m(\hat{k}_n < k_n^*(s_m) - \rho) = o_{\mathcal{M}_s}(1)$ can be shown similarly with the second statement of Lemma 3.23 such that the assertion follows. □

Proof of Lemma 3.23. Recall the statement of Lemma 3.21, (iii), and assume that n is such that for any $k \in \mathcal{K}_n$ it holds that $k \geq k_0$, where k_0 as defined in $\widetilde{\mathcal{M}}$, see (3.16). For sake of convenience we will write n for n_2 as this lemma depends only on the subsample S_2. Further, note if $m \in \widetilde{\mathcal{M}}$ and s_m as in Definition 3.6, then from the first term in (3.33) and (3.46) we have that

$$\frac{\widetilde{b_1}}{\widetilde{b_2}} B(r, s_m) \leq |\tilde{\theta}_{h_r,m} - \theta_m| \leq B(r, s_m), \tag{3.47}$$

by appropriate choice of C_{b_1}, C_{b_2} in (3.46), as the denominator in (3.33) is a non-vanishing constant which can be chosen uniformly in $\widetilde{\mathcal{M}}$ by Lemma 3.16 resp. Lemma 3.20.

Ad (i).

Fix some $k \in \mathcal{K}_n$ such that $k > k_n^*(s_m)$ and obtain by definition of $\hat{k}_n$ in (3.18)

$$P_m(\hat{k}_n = k) \leq \sum_{l \in \mathcal{K}_n : l \geq k} P_m\big(|\hat{\theta}_{h_{k-1},n} - \hat{\theta}_{h_l,n}| > C_{Lep}\sigma(n,l)\big). \tag{3.48}$$

From now on let $l \in \mathcal{K}_n$ such that $l \geq k$. Using (3.47) derive that

$$|\hat{\theta}_{h_{k-1},n} - \hat{\theta}_{h_l,n}| \leq |\hat{\theta}_{h_{k-1},n} - \tilde{\theta}_{h_{k-1},m}| + |\hat{\theta}_{h_l,n} - \tilde{\theta}_{h_l,m}| + B(k-1,s) + B(l,s).$$

By Lemma 3.21, (ii), and the definition of $k_n^*(s_m)$ obtain

$$B(k-1,s_m) + B(l,s_m) \leq 2B(k_n^*(s_m),s_m) \leq C_{Lep}\sigma\big(n,k_n^*(s_m)\big)/4 \leq C_{Lep}\sigma(n,l)/4.$$

Combining the two latter displays and Lemma 3.15 with $\tau = 1/3$ and $\bar{C} = C_{Lep}$ lead us for sufficiently large n to

$$\begin{aligned} & P_m\big(|\hat{\theta}_{h_{k-1},n} - \hat{\theta}_{h_l,n}| > C_{Lep}\sigma(n,l)\big) \\ & \quad \le P_m\Big(|\hat{\theta}_{h_{k-1},n} - \tilde{\theta}_{h_{k-1},m}| > \frac{3C_{Lep}}{8}\sigma(n,l)\Big) + P_m\Big(|\tilde{\theta}_{h_l,m} - \hat{\theta}_{h_l,n}| > \frac{3C_{Lep}}{8}\sigma(n,l)\Big) \\ & \quad \le 1_{\{|\theta_m - \tilde{\theta}_{h_{k-1},m}| > C_{Lep}\sigma(n,l)/8\}} + 2\exp\big(-C_1 C_{Lep}^2 n h_{k-1}^{2\gamma-1}\sigma^2(n,l)\big) \\ & \qquad + 1_{\{|\tilde{\theta}_{h_l,m} - \theta_m| > C_{Lep}\sigma(n,l)/8\}} + 2\exp\big(-C_2 C_{Lep}^2 n h_l^{2\gamma-1}\sigma^2(n,l)\big) \\ & \quad =: (A) + (B) + (C) + (D), \end{aligned} \tag{3.49}$$

for some absolute constants $C_i > 0$, $i = 1,2$ depending only on K,σ,σ_g as well as on L, Θ and s_m of $\widetilde{\mathcal{M}_s}$, as in (3.17). Since the constants C_1 and C_2 can be chosen continuously in s_m by Lemma 3.15 and $s_m \in [\underline{s},\overline{s}]$, we can choose these constants even uniformly in $\widetilde{\mathcal{M}}$. Using Lemma 3.21, (v), the terms (A) and (C) vanish for n large enough. Due to $l \le k$ we have by Lemma 3.21, (i), that $h_{k-1} > h_l$ such that by Lemma 3.21, (vi) and (vii), and if $C_{Lep} > 0$ is chosen large enough (depending only on K,σ,σ_g,L and Θ)

$$(B) \le \exp\Big(-C_1 C_{Lep}^2 n h_l^{2\gamma-1}\sigma^2(n,l)\Big) \le \exp\Big(-C_1 C_{Lep}^2 n h_l^{2\gamma+1}\Big) \le C_3\, 2^{-l/C_3},$$

for some finite constant $C_3 > 0$ depending only on K,σ,σ_g as well as on L and Θ of $\widetilde{\mathcal{M}}$. Similarly, with an appropriate choice of $C_{Lep} > 0$ (depending only on K,σ,σ_g,L and Θ) and Lemma 3.21, (vi), deduce that $(D) \le C_4\, 2^{-l/C_4}$, for some finite constant $C_4 > 0$ (depending only on K,σ,σ_g,L and Θ). Setting $c_1 = \max\{C_3,C_4\}$ yields (i) in view of (3.48), (3.49) and the analysis of the terms (A)–(D) as well as since C_3 and C_4 are uniform in $\widetilde{\mathcal{M}}$.

Ad (ii).

Fix some $k < k_n^*(s_m) - \rho$, where $\rho \in \mathbb{N}$ is chosen below. By definition of $\hat{k}_n$ in (3.18) and since $k < k_n^*(s_m)$,

$$P_m(\hat{k}_n = k) \le P_m\Big(|\hat{\theta}_{h_k,n} - \hat{\theta}_{h_{k_n^*(s_m)},n}| \le C_{Lep}\sigma\big(n,k_n^*(s_m)\big)\Big). \tag{3.50}$$

Further, by means of (3.47) and the reverse triangle inequality

$$\begin{aligned} & |\hat{\theta}_{h_k,n} - \hat{\theta}_{h_{k_n^*(s_m)},n}| \\ & \quad \ge \Big(\frac{\widetilde{b_1}}{\widetilde{b_2}}\Big) B(k,s_m) - B(k_n^*(s_m),s_m) - |\hat{\theta}_{h_k,n} - \tilde{\theta}_{h_k,m} - \hat{\theta}_{h_{k_n^*(s_m)},n} + \tilde{\theta}_{h_{k_n^*(s_m)},m}|. \end{aligned} \tag{3.51}$$

By using Lemma 3.21, (viii), twice as well as the fact that $k_n^*(s_m) - k > \rho$ and $s \ge \underline{s}$ yields

$$\begin{aligned} \Big(\frac{\widetilde{b_1}}{\widetilde{b_2}}\Big) B(k,s_m) - B(k_n^*(s_m),s_m) &= \Big(\frac{\widetilde{b_1}}{\widetilde{b_2}} 2^{(s-\gamma+1)(k_n^*(s_m)-k)} - 1\Big) B(k_n^*(s_m),s_m) \\ &= \Big(\frac{\widetilde{b_1}}{\widetilde{b_2}} 2^{(s-\gamma+1)(k_n^*(s_m)-k)} - 1\Big) 2^{-(s-\gamma+1)} B(k_n^*(s_m)-1,s_m) \\ &> \Big(\frac{\widetilde{b_1}}{\widetilde{b_2}} 2^{(s-\gamma+1)(\rho-1)} - 2^{-(s-\gamma+1)}\Big) B(k_n^*(s_m)-1,s_m) \\ &\ge \Big(\frac{\widetilde{b_1}}{\widetilde{b_2}} 2^{(\underline{s}-\gamma+1)(\rho-1)} - 2^{-(\underline{s}-\gamma+1)}\Big) B(k_n^*(s_m)-1,s_m). \end{aligned}$$

Next, by Lemma 3.21, (ii), (iv) and (ix),

$$
\begin{aligned}
B(k_n^*(s_m)-1,s_m) &\geq C_{Lep}\sigma(n,k_n^*(s_m)-1)/8 = C_{Lep}\,\sigma\big(n,k_n^*(s_m)\big)2^{\frac{-(2\gamma+5)}{2}}\left(\sqrt{1+k_n^*(s_m)^{-1}}\right)^{-1} \\
&\geq 2^{-(\gamma+3)}C_{Lep}\sigma\big(n,k_n^*(s_m)\big),
\end{aligned}
$$

where we used for the last inequality that due to $k_n^*(s_m)\geq 2$ we have that $\sqrt{(k_n^*(s_m)-1)/k_n^*(s_m)}\geq 2^{-1/2}$. Let

$$
\tilde{C} := 2^{-(\gamma+3)}\,C_{Lep}\Big(\frac{\widetilde{b_1}}{\widetilde{b_2}}2^{(\underline{s}-\gamma+1)(\rho-1)}-2^{-(\underline{s}-\gamma+1)}\Big),
$$

which can be made arbitrarily large by choosing ρ appropriately and depending only on $b_1,b_2,\underline{s}$ and on C_{Lep}. In view of (3.51) we have just shown that

$$
\begin{aligned}
&|\hat{\theta}_{h_k,n}-\hat{\theta}_{h_{k_n^*(s_m)},n}| \\
&\quad\geq \tilde{C}\sigma(n,k_n^*(s_m))-|\hat{\theta}_{h_k,n}-\tilde{\theta}_{h_k,m}-\hat{\theta}_{h_{k_n^*(s_m)},n}+\tilde{\theta}_{h_{k_n^*(s_m)},m}|.
\end{aligned}
\tag{3.52}
$$

Thus, using (3.52) to bound (3.50) yields

$$
\begin{aligned}
P_m(\hat{k}_n=k) \leq\ & P_m\Big(|\hat{\theta}_{h_k,n}-\tilde{\theta}_{h_k,m}|\geq(\tilde{C}-C_{Lep})\sigma\big(n,k_n^*(s_m)\big)/2\Big) \\
&+P_m\Big(\hat{\theta}_{h_{k_n^*(s_m)},n}+\tilde{\theta}_{h_{k_n^*(s_m)},m}|\geq(\tilde{C}-C_{Lep})\sigma\big(n,k_n^*(s_m)\big)/2\Big)
\end{aligned}
$$

and one can proceed similarly as in the first part for the term (3.49) by choosing $\tilde{C}$ suitable by choice of ρ. □

Adaptive coverage

For sake of brevity, define for $k\in\mathcal{K}_n$ the studentized random variables

$$
\begin{aligned}
\widetilde{S}_{\theta_m}(h_{k+u_n},n_1) &:= \frac{\hat{\theta}_{h_{k+u_n},n_1}-\tilde{\theta}_{h_{k+u_n},n_1}}{\tilde{w}_{n_1}^{loc}(h_{k+u_n})}, \\
\widetilde{S}_{[m^{(\gamma)}]}(h_{k+v_n},n_1) &:= \frac{\widehat{[m^{(\gamma)}]}_{h_{k+v_n}}-\widetilde{[m^{(\gamma)}]}_{h_{k+v_n}}}{\tilde{w}_{n_1}^{mag}(h_{k+v_n})}
\end{aligned}
\tag{3.53}
$$

where $\tilde{w}_n^{loc}$ and $\tilde{w}_n^{mag}$ as defined in (3.13). In the same spirit,

$$
\begin{aligned}
S_{\theta_m}(h_{k+u_n},n_1) &:= \frac{\hat{\theta}_{h_{k+u_n},n_1}-\theta_m}{\hat{w}_{n_1}^{loc}(h_{k+u_n})}, \\
S_{[m^{(\gamma)}]}(h_{k+v_n},n_1) &:= \frac{\widehat{[m^{(\gamma)}]}_{h_{k_n^*(s_m)+v_n},n_1}-[m^{(\gamma)}]}{\hat{w}_{n_1}^{mag}(h_{k+v_n})},
\end{aligned}
\tag{3.54}
$$

where $\hat{w}_n^{loc}$ as defined in (3.19) and $\hat{w}_n^{mag}$ as in (3.20). With the next lemma we will be able to verify the first assertion of Theorem 3.7.

Lemma 3.24. *Let ρ be as in Lemma 3.22, then for sufficiently large n the following statements hold true for any $k\in[k_n^*(s_m)-\rho,k_n^*(s_m)]$.*

(i) For any $\alpha \in (0,1)$

$$\liminf_n \inf_{m\in\widetilde{\mathcal{M}}} P_m\Big(\big|\widetilde{S}_{\theta_m}(h_{k+u_n},n_1)\big| \le q_{1-\alpha/2}\big(N(0,1)\big)\Big) = 1-\alpha,$$
$$\liminf_n \inf_{m\in\widetilde{\mathcal{M}}} P_m\Big(\big|\widetilde{S}_{[m^{(\gamma)}]}(h_{k+v_n},n_1)\big| \le q_{1-\alpha/2}\big(N(0,1)\big)\Big) = 1-\alpha,$$
$$\liminf_n \inf_{m\in\widetilde{\mathcal{M}}} P_m\Big(\max\big(\big|\widetilde{S}_{\theta_m}(h_{k+u_n},n_1)\big|, \big|\widetilde{S}_{[m^{(\gamma)}]}(h_{k+v_n},n_1)\big|\big) \le q_{1-\alpha/2}(W)\Big) = 1-\alpha,$$

where $q_{1-\alpha/2}(W)$ as defined in (3.21).

(ii) Uniformly in $\widetilde{\mathcal{M}}$ it holds that

$$\Big|\big|S_{\theta_m}(h_{k+u_n},n_1)\big| - \big|\widetilde{S}_{\theta_m}(h_{k+u_n},n_1)\big|\Big| = o_P(1),$$
$$\Big|\big|S_{[m^{(\gamma)}]}(h_{k+v_n},n_1)\big| - \big|\widetilde{S}_{[m^{(\gamma)}]}(h_{k+v_n},n_1)\big|\Big| = o_P(1).$$

Proof of Lemma 3.24. For sake of brevity we write n for n_1 as only the subsample S_1 is involved. Let $m \in \widetilde{\mathcal{M}}$ and s_m be as in Definition 3.6. First, we show that any bandwidth resolution h_{k+u_n} with $k \in [k_n^*(s_m)-\rho, k_n^*(s_m)]$ satisfies the bandwidth conditions of Theorem 3.5, that is to be more precisely

$$nh_{k+u_n}^{4s_m-2\gamma+1} \to 0, \quad \text{and} \quad nh_{k+u_n}^{2\gamma+1}\log(h_{k+u_n}^{-1})^{-1} \to \infty. \tag{3.55}$$

Note that by Assumption 3.4 we have that $h_{k+u_n} \cong h_k \log(n)^{-1/2\gamma-1}$. Further, ρ is a finite constant, uniform in $\widetilde{\mathcal{M}}$ so that

$$h_k \cong h_{k_n^*(s_m)}, \quad \forall k \in [k_n^*(s_m)-\rho, k_n^*(s_m)].$$

As a consequence, we only need to check (3.55) for $\check{h} := h_{k_n^*(s_m)}\log(n)^{-1/2\gamma-1}$. For this purpose note that by Assumption 3.3 (recall that we have set $n = n_1$) combined with Lemma 3.21, (iv), we have that

$$\check{h} \cong \frac{\log(n)^{\frac{2(\gamma-s_m-1)}{(2s_m+1)(2\gamma-1)}}}{n^{1/(2s_m+1)}}.$$

Compute that

$$n\check{h}^{4s_m-2\gamma+1} \cong \frac{\log(n)^{\zeta_1}}{n^{\zeta_2}},$$

where $\zeta_1 \in \mathbb{R}$ (not relevant) and $\zeta_2 = 2(s_m-\gamma)/(2s_m+1) > 0$ since $s_m \ge \underline{s} \ge \gamma+1$. Thus, both sides of the latter display tend to zero. Moreover,

$$n\check{h}^{2\gamma+1}\log(1/\check{h})^{-1/2} \cong \log(n)^{\zeta_3}\, n^{\zeta_4}$$

where $\zeta_3 \in \mathbb{R}$ and $\zeta_4 = 2(s_m-\gamma)/(2s_m+1) > 0$ such that both sides of the latter display tend to infinity. It is straightforward to check (3.55) for h_{k+u_n} replaced by h_{k+v_n} with $k \in [k_n^*(s_m)-\rho, k_n^*(s_m)]$.

Ad (i).

With the above shown asymptotics of the bandwidths, Theorem 3.5 immediately implies the first two statements, while the third follows by the uniform continuous mapping theorem C.11 in combination with Theorem 3.5. Indeed, note that $(x,y) \mapsto \max\{x,y\}$ is a continuous function and Φ does not depend on $\widetilde{\mathcal{M}}$, which trivially implies the condition of the uniform continuous mapping theorem C.11.

Ad (ii).

Let $k \in [k_n^*(s_m) - \rho, k_n^*(s_m)]$ be fixed. For convenience we write $\hat{\theta} := \hat{\theta}_{h_{k+u_n}, n_1}$ and $\tilde{\theta} := \tilde{\theta}_{h_{k+u_n}, n_1}$, as well as $\ddot{h} := h_{k+u_n}$, which has the same asymptotics as $\check{h}$ above. Recall the definitions in (3.53) and (3.54) and obtain

$$\begin{aligned}
\Big\| S_{\theta_m}(h_{k+u_n}, n_1)\big| - \big|\widetilde{S}_{\theta_m}(h_{k+u_n}, n_1)\Big\| \\
&= \Big\| \frac{\hat{\theta} - \theta_m}{\hat{w}_n^{loc}(\ddot{h})}\Big| - \Big|\frac{\hat{\theta} - \tilde{\theta}}{\tilde{w}_n^{loc}(\ddot{h})}\Big\| \\
&\le \Big|\Big(\frac{1}{\tilde{w}_n^{loc}(\ddot{h})} - \frac{1}{\hat{w}_n^{loc}(\ddot{h})}\Big)(\hat{\theta} - \theta_m)\Big| + \Big|\frac{\theta_m - \tilde{\theta}}{\tilde{w}_n^{loc}(\ddot{h})}\Big| \qquad (3.56)\\
&= \frac{|K^{(2)}(0)|}{\|K^{(\gamma+2)}\|_2}\Big[\Big|\Big(\frac{\widehat{[m^{(\gamma)}]}_{h_k}}{\hat{\sigma}} - \frac{[m^{(\gamma)}]}{\sigma}\Big)\sqrt{n\ddot{h}^{2\gamma-1}}(\hat{\theta} - \theta_m)\Big| \\
&\quad + \Big|\sqrt{n\ddot{h}^{2\gamma-1}}\,\frac{(\theta_m - \tilde{\theta})[m^{(\gamma)}]}{\sigma}\Big|\Big],
\end{aligned}$$

where the last line is due to definition of $\tilde{w}_n^{loc}$ resp. $\hat{w}_n^{loc}$ in (3.13) resp. (3.19). For the following argumentation note that the constants in the O-terms of Theorem 3.3 and 3.4 depend only on K, σ_g, L as well as s_m and are continuous in s_m. Thus, since $s_m \in [\underline{s}, \overline{s}]$ the constants in these O-terms hold uniformly over $\widetilde{\mathcal{M}}$ as well. As a consequence, by Theorem 3.4 and the assumption on $\hat{\sigma}$ in Theorem 3.7 we have that

$$\Big|\frac{\widehat{[m^{(\gamma)}]}_{h_k}}{\hat{\sigma}} - \frac{[m^{(\gamma)}]}{\sigma}\Big| = O_{P,\widetilde{\mathcal{M}}}(h_k^{s_m-\gamma}) + O_{P,\widetilde{\mathcal{M}}}\big((nh_k^{2\gamma+1})^{-1/2}\big).$$

This together with Theorem 3.3 implies that the first term on the right-hand side of (3.56) is $o_{P,\widetilde{\mathcal{M}}}(1)$. Moreover, Theorem 3.3 shows that the second term is of order

$$O_{\widetilde{\mathcal{M}}}(\sqrt{n}\ddot{h}^{s_m+1/2}) = o_{P,\widetilde{\mathcal{M}}}(1),$$

due to the undersmoothing, which concludes the first claim of (ii). The second assertion can be dealt with similarly. □

Proof of (3.22). Let ρ be as in Lemma 3.22 and n large enough such that the statements of Lemma 3.22 and Lemma 3.24 hold. Then, for any $\alpha \in (0,1)$

$$\begin{aligned}
P_m&\Big((\theta_m, [m^{(\gamma)}])^T \in C_n^{loc}(\alpha) \times C_n^{mag}(\alpha)\Big) \\
&= P_m\Big(\big(|S_{\theta_m}(h_{\hat{k}_n+u_n}, n_1)|, |S_{[m^{(\gamma)}]}(h_{\hat{k}_n+v_n}, n_1)|\big)^T \le q_{1-\alpha/2}(W)(1,1)^T\Big) \\
&= \sum_{k \in \mathcal{K}_n} P_m\Big(\big(|S_{\theta_m}(h_{k+u_n}, n_1)|, |S_{[m^{(\gamma)}]}(h_{k+v_n}, n_1)|\big)^T \le q_{1-\alpha/2}(W)(1,1)^T, \{\hat{k}_n = k\}\Big) \\
&= \sum_{k_n^*(s_m)-\rho \le k \le k_n^*(s_m)} P_m\Big(\big(|S_{\theta_m}(h_{k+u_n}, n_1)|, |S_{[m^{(\gamma)}]}(h_{k+v_n}, n_1)|\big)^T \le q_{1-\alpha/2}(W)(1,1)^T, \{\hat{k}_n = k\}\Big) \\
&\quad + o_{P,\widetilde{\mathcal{M}}}(1), \\
&= \sum_{k_n^*(s_m)-\rho \le k \le k_n^*(s_m)} P_m\Big(\big(|\widetilde{S}_{\theta_m}(h_{k+u_n}, n_1)|, |\widetilde{S}_{[m^{(\gamma)}]}(h_{k+v_n}, n_1)|\big)^T \le q_{1-\alpha/2}(W)(1,1)^T, \{\hat{k}_n = k\}\Big) \\
&\quad + o_{P,\widetilde{\mathcal{M}}}(1),
\end{aligned}$$

where we used Lemma 3.22 for the third equality and Lemma 3.24, (ii), in combination with Theorem C.5 for the last equality. Notice that $\widetilde{S}_{\theta_m}(h_{k+u_n},n_1)$ and $\widetilde{S}_{[m^{(\gamma)}]}(h_{k+v_n},n_1)$ are independent of $\hat{k}_n$ since they are based on the different subsamples S_1 resp. S_2 in Assumption 3.3. Thus,

$$\begin{aligned}
&P_m\big((\theta_m,[m^{(\gamma)}])^T \in C_{n,f}(\alpha/2)\times C_{n,[m^{(\gamma)}]}(\alpha/2)\big)\\
&\quad\geq (1-\alpha)\sum_{k_n^*(s_m)-\rho\leq k\leq k_n^*(s_m)} P_m(\hat{k}_n=k)\\
&\quad+\sum_{k_n^*(s_m)-\rho\leq k\leq k_n^*(s_m)}\Big(P_m\big(\max\big(\big|\widetilde{S}_{\theta_m}(h_{k+u_n},n_1)\big|,\big|\widetilde{S}_{[m^{(\gamma)}]}(h_{k+v_n},n_1)\big|\big)^T\leq q_{1-\alpha/2}(W)\big)-(1-\alpha)\Big)P_m(\hat{k}_n=k)\\
&\quad+o_{P,\widetilde{\mathcal{M}}}(1).
\end{aligned}$$

The first term converges to $1-\alpha$ by means of Lemma 3.22, while the second term is asymptotically greater or equal zero by Lemma 3.24, (i). This concludes the proof of (3.22). □

Adaptive length

The next lemma suffices to prove (3.23), the adaptive length of the confidence sets.

Lemma 3.25. *There exist finite constants $\tilde{C}_1,\tilde{C}_2>0$ uniform in $\widetilde{\mathcal{M}}$ such that*

$$P_m\Big(1/\sqrt{nh^{2\gamma-1}_{\hat{k}_n+u_n}}\geq\tilde{C}_1\left(\log(n)/n\right)^{(s_m-\gamma+1)/(2s_m+1)}\Big)=o_{P,\widetilde{\mathcal{M}}}(1),$$
$$P_m\Big(1/\sqrt{nh^{2\gamma+1}_{\hat{k}_n+v_n}}\geq\tilde{C}_2\left(\log(n)/n\right)^{(s_m-\gamma)/(2s_m+1)}\Big)=o_{P,\widetilde{\mathcal{M}}}(1),$$

where s_m as in Definition 3.6.

Proof of Lemma 3.25. Let $0<\tilde{C}\leq 1$. By Assumption 3.4 and Lemma 3.21, (iv) and (i), obtain

$$\begin{aligned}
&\sup_{m\in\widetilde{\mathcal{M}}} P_m\Big(1/\sqrt{nh^{2\gamma-1}_{\hat{k}_n+u_n}}\geq\tilde{C}_1\left(\log(n)/n\right)^{(s_m-\gamma+1)/(2s+1)}\Big)\\
&\qquad\leq\sup_{m\in\widetilde{\mathcal{M}}} P_m\Big(h^{-2\gamma+1}_{\hat{k}_n+u_n}\geq n\left(\log(n)/n\right)^{\frac{2(s_m-\gamma+1)}{2s_m+1}}\Big)\\
&\qquad=\sup_{m\in\widetilde{\mathcal{M}}} P_m\Big(h^{-1}_{\hat{k}_n}\geq\left(\frac{n}{\log(n)}\right)^{\frac{1}{2\gamma-1}}\left(\frac{\log(n)}{n}\right)^{\frac{2(s_m-\gamma+1)}{(2s_m+1)(2\gamma-1)}}\Big)\\
&\qquad=\sup_{m\in\widetilde{\mathcal{M}}} P_m(h^{-1}_{\hat{k}_n}\geq h^{-1}_{k_n^*(s_m)})\\
&\qquad=\sup_{m\in\widetilde{\mathcal{M}}} P_m(\hat{k}_n\geq k_n^*(s_m))=o(1),
\end{aligned}$$

where the last equality is due to Lemma 3.22. The second assertion can be derived similarly. □

Proof of (3.23). Note that

$$\begin{pmatrix}\hat{w}_n^{loc}\\ \hat{w}_n^{mag}\end{pmatrix}=\frac{\hat{\sigma}}{K^{(2)}(0)}\begin{pmatrix}1/\widehat{[m^{(\gamma)}]}_{h_{\hat{k}_n},n_1}\\ 1\end{pmatrix}\cdot\begin{pmatrix}1/\sqrt{nh^{2\gamma-1}_{\hat{k}_n+u_n}}\\ 1/\sqrt{nh^{2\gamma+1}_{\hat{k}_n+v_n}}\end{pmatrix},$$

where the multiplication is to be understood componentwise. Now the first vector on the right-hand side of the latter display converges uniformly over $\widetilde{\mathcal{M}}$ in distribution, due to assumption on $\hat{\sigma}$ and

(3.22), which implies the uniform convergence of $\widehat{[m^{(\gamma)}]}_{h_{\hat{k}_n},n_1}$ in distribution. Concerning the second term, Lemma 3.25 shows

$$\limsup_n \sup_{m\in\widetilde{\mathcal{M}}} P_m\left(\begin{pmatrix} 1/\sqrt{nh^{2\gamma-1}_{\hat{k}_n+u_n}} \\ 1/\sqrt{nh^{2\gamma+1}_{\hat{k}_n+v_n}} \end{pmatrix} \geq C\left(\log(n)/n\right)^{(s-\gamma)/(2s+1)} \begin{pmatrix} \left(\log(n)/n\right)^{1/(2s+1)} \\ 1 \end{pmatrix}\right) = 0,$$

where $C = \max\{\tilde{C}_1, \tilde{C}_2\}$ and $\tilde{C}_i$ as in Lemma 3.25 for $i = 1,2$ respectively. Thus, conclude the proof by the uniform Slutzky Theorem C.12. ◻

3.6 Kernel construction

In this section the kernel construction of Cheng and Raimondo (2008) will be adapted in order to construct kernels which satisfy Assumption 3.2 for $\gamma = 1$ and $s \geq \gamma + 1$.

Remark 11. Assumption 3.2 is similar to the kernel assumption stated in Cheng and Raimondo (2008) which are listed in their assumption $C_{\alpha,s}$, but stricter as we see in the following. Indeed, a comparison of both kernel assumptions reveals that we claim $(\gamma+5)$-times continuous differentiability of K, instead of $(\gamma+3)$-times, due to the Taylor expansions in (3.33) or (3.35). However, their constructed kernels are infinitely often differentiable, so this assumption is satisfied for their suggested kernels. The critical point is that we demand more derivatives of K to vanish at the boundary points (for the analysis in Lemma 3.8), which is not implied by $C_{\alpha,s}$ and not satisfied by their constructed kernels. As a result, their constructed kernels can not be used for our results.

Preliminaries for the construction

Note that kernels fulfilling Assumption 3.2 for s such that $\lfloor s \rfloor$ is odd also fulfill the assumption for $s+1$. Hence, we only construct kernels for $\lfloor s \rfloor$ odd. In particular, we will construct the kernel K by its second derivative $K^{(2)}$. For the construction of $K^{(2)}$ we restrict to polynomials of even degrees up to $\lfloor s \rfloor + 5$.

Let $\lfloor s \rfloor$ be odd and $k \in \mathbb{N}$ such that $\lfloor s \rfloor = 2k+1$ and let P_k denote the k-th Legendre polynomial, that is

$$\mathsf{P}_k(x) = 2^{-k} \sum_{j=0}^{\lfloor k/2 \rfloor} (-1)^j \frac{(2k-2j)!}{j!(k-j)!(k-2j)!} x^{n-2j} 1_{[-1,1]}(x).$$

By Rodriguéz formula the latter expression can be rewritten to

$$\mathsf{P}_k(x) = 1_{[-1,1]}(x) \frac{1}{2^k k!} \frac{\partial^k}{\partial x^k}(x^2-1)^k. \tag{3.57}$$

Since we restrict to polynomials of even degrees up to $\lfloor s \rfloor + 5$, which equals $2(k+3)$, we can write

$$K^{(2)}(x) = \sum_{i=0}^{k+3} p_{2i}\, \mathsf{P}_{2i}(x), \tag{3.58}$$

due to orthogonality of $(\mathsf{P}_k)_k$, where

$$p_i := \frac{\int_{-1}^{1} \mathsf{P}_i(x) K^{(2)}(x)}{||\mathsf{P}_i||^2}.$$

Note that $||\mathsf{P}_i||^2 = 2/(2i+1)$ and that

$$\int_{-1}^{1} \mathsf{P}_{2k} K^{(2)}(x)\,\mathrm{d}x = 2^{-2k}\sum_{i=0}^{k}(-1)^i \frac{(4k-2i)!}{i!(2k-i)!(2k-2i)!}\int_{-1}^{1} x^{2(k-i)}K^{(2)}(x)\,\mathrm{d}x$$
$$= 2^{-2k}\sum_{i=0}^{k}(-1)^i a(2k,i)\,M_{2k-2i},$$

where

$$a(k,i) = (-1)^i \frac{(2k-2i)!}{i!(k-i)!(k-2i)!}, \quad \text{and} \quad M_k = \int_{-1}^{1} x^k K^{(2)}(x)\,\mathrm{d}x.$$

In the following, we assume that $M_j = 0$ for $j = 0,1,\ldots,2k-1$ and compute the coefficients such that this property and the other conditions of Assumption 3.2 are satisfied. Hence, by assumption on K we have that $M_j = 0$ for $j = 0,1,\ldots,2k-1$, so that

$$p_i = \begin{cases} 0, & i < 2k \text{ or } i \text{ is odd} \\ \frac{4k+1}{2} a(2k,0)M_{2k}, & i = 2k \\ \frac{4k+5}{2}\big(a(2k+2,0)M_{2k+2} - a(2k+2,1)M_{2k}\big), & i = 2k+2 \\ \frac{4k+9}{2}\big(a(2k+4,0)M_{2k+4} - a(2k+4,1)M_{2k+2} + a(2k+4,0)M_{2k}\big), & i = 2k+4 \\ \frac{4k+13}{2}\big(a(2k+6,0)M_{2k+6} - a(2k+6,1)M_{2k+4} + a(2k+6,0)M_{2k+2} - a(2k+6,0)M_{2k}\big), & i = 2k+6. \end{cases}$$

Consequently, $K^{(2)}$ in (3.58) reduces to

$$K^{(2)}(x) = \sum_{j=1}^{4} M_{2(k+j-1)}\tilde{\mathsf{P}}_j(k,x), \tag{3.59}$$

where

$$\begin{aligned}
\tilde{\mathsf{P}}_1(k,x) &= \frac{4k+1}{2}a(2k,0)\mathsf{P}_{2k}(x) - \frac{4k+5}{2}a(2k+2,1)\mathsf{P}_{2k+2}(x) \\
&\quad + \frac{4k+9}{2}a(2k+4,2)\mathsf{P}_{2k+4}(x) - \frac{4k+13}{2}a(2k+6,1)\mathsf{P}_{2k+6}(x), \\
\tilde{\mathsf{P}}_2(k,x) &= \frac{4k+5}{2}a(2k+2,0)\mathsf{P}_{2k+2}(x) - \frac{4k+9}{2}a(2k+4,1)\mathsf{P}_{2k+4}(x) \\
&\quad + \frac{4k+13}{2}a(2k+6,2)\mathsf{P}_{2k+6}(x), \\
\tilde{\mathsf{P}}_3(k,x) &= \frac{4k+9}{2}a(2k+4,0)\mathsf{P}_{2k+4}(x) - \frac{4k+13}{2}a(2k+6,1)\mathsf{P}_{2k+6}(x), \\
\tilde{\mathsf{P}}_4(k,x) &= \frac{4k+13}{2}a(2k+6,0)\mathsf{P}_{2k+6}(x).
\end{aligned}$$

Vanishing edges

First, we require $K^{(2)}(1) = 0$. By properties of the Legendre polynomials it holds that $\mathsf{P}_n(1) = 1$ for every $n \in \mathbb{N}$. Thus, by simplification

$$\begin{aligned}
\tilde{\mathsf{P}}_1(k,1) &= \frac{(4k+1)!}{2^{2k+1}((2k)!)^2} - \frac{(4k+5)(4k+2)!}{2^{2k+3}(2k+1)!(2k)!} + \frac{(4k+9)(4k+4)!}{2^{2k+6}(2k+2)!(2k)!} - \frac{(4k+13)(4k+6)!}{3!2^{2k+7}(2k+3)!(2k)!} \\
&= \frac{(4k+1)!}{2^{2k+1}((2k)!)^2}\Big(1 - \frac{(4k+5)}{2} + \frac{(4k+3)(4k+9)}{2^3} - \frac{(4k+3)(4k+5)(4k+13)}{3\cdot 2^4}\Big) \\
&= -\frac{(4k+1)!}{3\cdot 2^{2k+5}((2k)!)^2}(4k+3)(4k+5)(4k+7).
\end{aligned}$$

Similarly,

$$
\begin{aligned}
\tilde{\mathsf{P}}_2(k,1) &= \frac{(4k+5)!}{2^{2k+6}((2k+2)!)^2}(4k+7)(4k+9),\\
\tilde{\mathsf{P}}_3(k,1) &= -\frac{(4k+9)!}{2^{2k+6}((2k+4)!)^2}(4k+11),\\
\tilde{\mathsf{P}}_4(k,1) &= \frac{(4k+13)!}{2^{2k+7}((2k+6)!)^2}.
\end{aligned}
$$

Since we claim that $K^{(2)}(1)=0$, we obtain by (3.59) and the two latter displays that it should hold

$$
M_{2k} = \frac{3\cdot 2(4k+9)}{(2k+1)(2k+2)}\Big(M_{2k+2} - \frac{2^2(4k+11)}{(2k+3)(2k+4)} + \frac{2^3(4k+11)(4k+13)}{(2k+3)(2k+4)(2k+5)(2k+6)}\Big). \tag{3.60}
$$

Note that $K^{(2)}(-1)=0$ also holds in this case, since $K^{(2)}$ is even.

Next, we require $K^{(3)}(1)=0$. From (3.59) the third derivative of K is

$$
K^{(3)}(x) = \sum_{j=1}^{4} M_{2(k+j-1)}\tilde{\mathsf{P}}_j^{(1)}(k,x).
$$

Now, recall that for the derivatives of Legendre polynomials one has $\mathsf{P}_n^{(1)}(1) = {}^{n(n+1)}/_2$. Thus,

$$
\begin{aligned}
\tilde{\mathsf{P}}_1^{(1)}(k,1) &= \frac{(4k+1)!}{2^{2k+2}((2k)!)^2}2k(2k+1) - \frac{(4k+5)(4k+2)!}{2^{2k+4}(2k+1)!(2k)!}(2k+2)(2k+3)\\
&\quad + \frac{(4k+9)(4k+4)!}{2^{2k+6}(2k+2)!(2k)!}(2k+4)(2k+5) - \frac{(4k+13)(4k+6)!}{2^{2k+8}3!(2k+3)!(2k)!}(2k+6)(2k+7)\\
&= -\frac{(4k+1)!}{3\cdot 2^{2k+5}((2k)!)^2}(4k+3)(4k+5)(4k+7)(2k^2+13k+27).
\end{aligned}
$$

Likewise,

$$
\begin{aligned}
\tilde{\mathsf{P}}_2^{(1)}(k,1) &= \frac{(4k+5)!}{2^{2k+6}((2k+2)!)^2}(4k+7)(4k+9)(2k^2+13k+25),\\
\tilde{\mathsf{P}}_3^{(1)}(k,1) &= -\frac{(4k+9)!}{2^{2k+6}((2k+4)!)^2}(4k+11)(2k^2+13k+23),\\
\tilde{\mathsf{P}}_4^{(1)}(k,1) &= \frac{(4k+13)!}{2^{2k+7}((2k+6)!)^2}(2k^2+13k+21).
\end{aligned}
$$

For $K^{(3)}(1)=0$ to hold, it must hold from the latter three displays that

$$
\begin{aligned}
M_{2k} = b_1\frac{3\cdot 2(4k+9)}{(2k+1)(2k+2)}\Big(&M_{2k+2} - b_2\frac{2^2(4k+11)}{(2k+3)(2k+4)}M_{2k+4}\\
&+ b_3\frac{2^3(4k+11)(4k+13)}{(2k+3)(2k+4)(2k+5)(2k+6)}M_{2k+6}\Big),
\end{aligned} \tag{3.61}
$$

where

$$
b_1 = \frac{2k^2+13k+25}{2k^2+13k+27}, \quad b_2 = \frac{2k^2+13k+23}{2k^2+13k+25}, \quad b_3 = \frac{2k^2+13k+21}{2k^2+13k+25}.
$$

Since $K^{(3)}$ is odd this implies $K^{(3)}(-1)=0$.

Finally, we also require that $K^{(4)}(1) = 0$. From (3.59) obtain that

$$K^{(4)}(x) = \sum_{j=1}^{4} M_{2(k+j-1)} \tilde{\mathsf{P}}_j^{(2)}(k,x).$$

For the second derivatives of the Legendre polynomials one has $\mathsf{P}_n^{(2)}(1) = {}^{(n-1)n(n+1)(n+2)}/_{2^3}$. Indeed, using representation (3.57),

$$\begin{aligned}
\frac{\partial^2 \mathsf{P}_{2l}(x)}{\partial^2 x}\Big|_{x=1} &= \frac{1}{2^{2l}(2l)!}\frac{\partial^{2l+2}}{\partial x^{2l+2}}\left((x+1)^{2l}(x-1)^{2l}\right)\Big|_{x=1} \\
&= \frac{1}{2^{2l}(2l)!}\sum_{j=2}^{2l+2}\binom{2l+2}{j}\left[\frac{\partial^j}{\partial x^j}(x+1)^{2l}\,\frac{\partial^{2l+2-j}}{\partial x^{2l+2-j}}(x-1)^{2l}\right]_{x=1} \\
&= \frac{1}{2^{2l}(2l)!}\sum_{j=2}^{2l+2}\binom{2l+2}{j}\frac{(2l)!}{(2l-j)!}(x+1)^{2l-j}\,\frac{(2l)!}{(j-2)!}(x-1)^{j-2}\Big|_{x=1} \\
&= \frac{(2l-1)2l(2l+1)(2l+2)}{2^3},
\end{aligned}$$

since the sum is only for $j = 2$ not equal to zero. With this,

$$\begin{aligned}
\tilde{\mathsf{P}}_1^{(2)}(k,1) &= -\frac{(4k+1)!}{3\cdot 2^{2k+6}((2k)!)^2}(4k+3)(4k+5)(4k+7)(4k^4+52k^3+275k^2+665k+648), \\
\tilde{\mathsf{P}}_2^{(2)}(k,1) &= \frac{(4k+5)!}{2^{2k+7}((2k+2)!)^2}(4k+7)(4k+9)(4k^4+52k^3+267k^2+621k+556), \\
\tilde{\mathsf{P}}_3^{(2)}(k,1) &= -\frac{(4k+9)!}{2^{2k+7}((2k+4)!)^2}(4k+11)(4k^4+52k^3+259k^2+577k+480), \\
\tilde{\mathsf{P}}_4^{(2)}(k,1) &= \frac{(4k+13)!}{2^{2k+8}((2k+6)!)^2}(4k^4+52k^3+251k^2+533k+420).
\end{aligned}$$

So for $K^{(4)}(1) = 0$ to hold, it is required that

$$\begin{aligned}
M_{2k} = c_1\frac{3\cdot 2(4k+9)}{(2k+1)(2k+2)}\Big(M_{2k+2} - c_2\frac{2^2(4k+11)}{(2k+3)(2k+4)}M_{2k+4} \\
+ c_3\frac{2^3(4k+11)(4k+13)}{(2k+3)(2k+4)(2k+5)(2k+6)}M_{2k+6}\Big),
\end{aligned} \tag{3.62}$$

where

$$\begin{aligned}
c_1 &= \frac{4k^4+52k^3+267k^2+621k+556}{4k^4+52k^3+275k^2+665k+648}, \quad c_2 = \frac{4k^4+52k^3+259k^2+577k+480}{4k^4+52k^3+267k^2+621k+556} \\
c_3 &= \frac{4k^4+52k^3+251k^2+533k+420}{4k^4+52k^3+267k^2+621k+556}.
\end{aligned}$$

Since $K^{(4)}$ is even it follows that $K^{(4)}(-1) = 0$ as well.

Explicit construction

Summarizing the conditions in (3.60) – (3.62), we need to solve $\mathbf{Ax} = \mathbf{y}$, where

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} \frac{3\cdot 2(4k+9)}{(2k+1)(2k+2)} M_{2k+2} \\ -\frac{3\cdot 2^3(4k+9)(4k+11)}{(2k+1)(2k+2)(2k+3)(2k+4)} M_{2k+4} \\ \frac{3\cdot 2^4(4k+9)(4k+11)(4k+13)}{(2k+1)(2k+2)(2k+3)(2k+4)(2k+5)/2k+6} M_{2k+6} \end{pmatrix}, \mathbf{y} = \begin{pmatrix} M_{2k} \\ M_{2k} \\ M_{2k} \end{pmatrix}.$$

Now, for the inverse matrix of $\mathbf{A}$ it holds that $\mathbf{A}^{-1}(1,1,1)^T = (3,-3,1)^T$, so that

$$\begin{pmatrix} M_{2k+2} \\ M_{2k+4} \\ M_{2k+6} \end{pmatrix} = \frac{(2k+1)(2k+2)}{2(4k+9)} \begin{pmatrix} 1 \\ \frac{(2k+3)(2k+4)}{2^2(4k+11)} \\ \frac{(2k+3)(2k+4)(2k+5)(2k+6)}{3\cdot 2^3(4k+11)(4k+13)} \end{pmatrix} M_{2k}.$$

Plugging this into (3.59) and ordering the coefficients of the Legendre polynomials, we obtain that the coefficient of P_{2k+6} is

$$\begin{aligned}
&\frac{(4k+13)!}{2^{2k+7}((2k+6)!)^2} M_{2k+6} - \frac{(4k+13)(4k+10)!}{2^{2k+7}(2k+4)!(2k+5)!} M_{2k+4} \\
&\quad + \frac{(4k+13)(4k+8)!}{2^{2k+8}(2k+4)!(2k+2)!} M_{2k+2} - \frac{(4k+13)(4k+6)!}{3\cdot 2^{2k+8}(2k+3)!(2k)!} M_{2k} \\
&= \frac{(4k+1)!}{2^{2k+1}((2k)!)^2} \Big(\frac{(4k+3)(4k+5)(4k+7)}{3\cdot 2^4} - \frac{(4k+3)(4k+5)(4k+7)(4k+13)}{2^4(4k+11)} \\
&\qquad\qquad + \frac{(4k+3)(4k+5)(4k+7)(4k+13)}{2^4(4k+9)} - \frac{(4k+3)(4k+5)(4k+13)}{2^4} \Big) M_{2k} \\
&= -\frac{(4k+1)!}{2^{2k+1}((2k)!)^2} \frac{(4k+3)(4k+5)}{(4k+9)(4k+11)} M_{2k}.
\end{aligned}$$

Similarly, the coefficients for P_{2k+4}, P_{2k+2} and P_{2k} are respectively

$$\frac{(4k+1)!}{2^{2k+1}((2k)!)^2} \frac{3(4k+3)}{4k+11} M_{2k}, \quad -\frac{(4k+1)!}{2^{2k+1}((2k)!)^2} \frac{3(4k+5)}{4k+9} M_{2k}, \quad \frac{(4k+1)!}{2^{2k+1}((2k)!)^2} M_{2k}.$$

Thus, obtain that

$$\begin{aligned}
K^{(2)}(x) = \frac{(4k+1)!}{2^{2k+1}((2k)!)^2} M_{2k} \Big(&\mathsf{P}_{2k}(x) - \frac{3(4k+5)}{4k+9} \mathsf{P}_{2k+2}(x) \\
&+ \frac{3(4k+3)}{4k+11} \mathsf{P}_{2k+4}(x) - \frac{(4k+3)(4k+5)}{(4k+9)(4k+11)} \mathsf{P}_{2k+6}(x) \Big).
\end{aligned}$$

Setting $M_{2k} = (-1)^{k+1}$ the preceding display can be simplified to

$$K^{(2)}(x) = \frac{(2k+1)(2k+2)(4k+3)(4k+5)(4k+7)}{2^{2k+4}(4k+9)} \sum_{i=0}^{k+3} (-1)^{k-i} \frac{(2(k+i))!}{(k+3-i)!(k+i)!(2i)!} x^{2i}.$$

By choice of k, i.e. $k = (\lfloor s \rfloor - 1)/2$, this is equivalent to

$$\begin{aligned}
K^{(2)}(x) = &\frac{\lfloor s \rfloor(\lfloor s \rfloor + 1)(2\lfloor s \rfloor + 1)(2\lfloor s \rfloor + 3)(2\lfloor s \rfloor + 5)}{2^{\lfloor s \rfloor + 3}(2\lfloor s \rfloor + 7)} \\
&\times \sum_{i=0}^{k+3} (-1)^{k-i} \frac{(\lfloor s \rfloor + 2i - 1)!}{((\lfloor s \rfloor + 5)/2)!((\lfloor s \rfloor - 1 + 2i)/2)!(2i)!} x^{2i}.
\end{aligned}$$

Discussion of the requirements

It follows that K with a second derivative as in the preceding display satisfies the Assumption 3.2 for $\gamma = 1$ and $s \geq \gamma + 1$. Indeed, the support of the Legendre polynomials is $[-1,1]$, so is the support of K. Furthermore, the Legendre polynomials are infinitely often continuous differentiable inside its support, which implies K to be infinitely often continuous differentiable. Furthermore,

Ad (ii). $K^{(2)}$ is an even function, as it consists of even Legendre polynomials. Thus, $K^{(1)}$ is odd.

Ad (i). By construction $K^{(j)}(-1) = K^{(j)}(1) = 0$ for $j = 2,3,4$. Since $K^{(1)}$ is odd it holds $K^{(1)}(0) = 0$. Moreover, $0 = M_0 = \int_{-1}^{1} K^{(2)} = 2K^{(1)}(1)$ so that $K^{(1)}(1) = K^{(1)}(-1) = 0$.

Ad (iii). Further, by construction $M_j = 0$ for $j = 1,\ldots,\lfloor s \rfloor - 2$ and $M_{2k} \neq 0$. Hence, integration by parts implies $\int_{-1}^{1} x^j K^{(1)}(x)\,\mathrm{d}x = 0$ for $j = 1,\ldots,\lfloor s \rfloor - 3$ and $\int_{-1}^{1} x^{\lfloor s \rfloor - 2} K^{(1)}(x)\,\mathrm{d}x \neq 0$, due to $M_{2k} \neq 0$.

Ad (iv). $K^{(2)}$ has a unique global minimum at zero with some value strictly smaller than zero (This is implied by the choice of M_{2k} as above). In addition, $K^{(2)}(1) = K^{(2)}(-1) = 0$ such that $K^{(1)}$ has a global maximum left from the origin and in addition there exists some $q_l \in (0,1)$ such that $K^{(1)}$ is strictly positive on $(-q_l, 0)$ and the global maximum is inside this interval.

Ad (v). As $K^{(2)}(0) \neq 0$ we can find $x^* \in (0,1)$ and $c_2 > 0$ such that $|K^{(1)}(x)| \geq c_2|x|$ for any $x \in [-x^*, x^*]$.

CHAPTER 4

Lower bounds for change-point-locations in nonparametric regression

This chapter is devoted to deriving minimax rates for the change-point regression problems considered in the chapters before. Section 4.1 provides a general theorem to obtain minimax lower bounds, which is crucial for the remaining parts of this chapter. The lower bounds in Section 4.2 are well-known, see Korostelev and Tsybakov (1993), while the lower bounds in Section 4.3 are new.

4.1 General theorem for deriving minimax lower bounds

In order to derive the optimal rate as in (1.7) the classical approach is as follows. First, the supremum in (1.6) is restricted to a finite number of hypotheses, that is for $M \in \mathbb{N}$

$$\inf_{T_n} \sup_{f \in \mathcal{F}} \mathbb{E}_f \big(w(r_n^{-1} d(T_n, f)) \big) \geq \inf_{T_n} \max_{f \in \{f_0, \ldots, f_M\}} \mathbb{E}_f \big(w(r_n^{-1} d(T_n, f)) \big),$$

where $\{f_0, \ldots, f_M\}$ is some finite set in $\mathcal{F}$. Second, by the Markov inequality the risk is bounded below by the probability-risk:

$$\mathbb{E}_f \big(w(r_n^{-1} d(\hat{f}_n, f)) \big) \geq w(A) P_f (r_n^{-1} d(\hat{f}_n, f) \geq A) = w(A) P_f (d(\hat{f}_n, f) \geq s_n),$$

where $A > 0$ and $s_n = A r_n$. Third, by an appropriate choice of the elements $\{f_0, \ldots, f_M\}$ the minimax estimation lower bound reduces to a lower bound of specific testing problems. Indeed, suppose that

$$d(f_i, f_j) \geq 2 s_n, \quad \forall 0 \leq i < j \leq M, \tag{4.1}$$

and let $\hat{\Psi}_n : \mathcal{Y}^{(n)} \to \{0, \ldots, M\}$ be the *minimum distance test* given by

$$d(\hat{f}_n, f_{\hat{\Psi}_n}) = \min_{m = 0, \ldots, M} d(\hat{f}_n, f_m),$$

where in case of ties one may choose any of the minimizers, and $(\mathcal{Y}^{(n)}, \mathcal{A}^{(n)})$ is the measurable space associated with the data. Then,

$$P_{f_i}(d(\hat{f}_n, f_i) \geq s_n) \geq P_{f_i}(\hat{\Psi}_n \neq i), \quad i = 0, 1, \ldots, M.$$

Finally, if we can find elements $\{f_0,\dots,f_M\}$ satisfying (4.1) the minimax risk can be bounded as follows

$$\inf_{T_n}\sup_{f\in\mathcal{F}}\mathbb{E}_f\left(w(r_n^{-1}d(T_n,f))\right)\geq\inf_{\Psi_n}\max_{0\leq i\leq M}P_{f_i}(\Psi_n\neq i),\tag{4.2}$$

where the infimum is taken over all tests Ψ_n. The right-hand side of (4.2) can be controlled by information theoretic tools such as the *Kullback-Leibler-distance* between two probability measures P and Q (both absolute continuous to some σ-finite measure ν) defined by

$$\mathbf{KL}(P,Q)=\begin{cases}\int\log\frac{dP}{dQ}\,dP, & \text{if } P\ll Q,\\ +\infty, & \text{else.}\end{cases}$$

A comprehensive account on this issue can be found in Tsybakov (2009). The following proposition is a modification of Proposition 2.3 in Tsybakov (2009) giving a lower bound for the right-hand side of (4.2).

Proposition 4.1. *Let $M\in\mathbb{N}$ and $P_0,\dots,P_M$ be probability measures with $P_i\ll P_0$ for $i=1,\dots,M$ and*

$$\frac{1}{M}\sum_{i=1}^{M}\boldsymbol{KL}(P_i,P_0)\leq\alpha\log(M),$$

for some $\alpha\in(0,1/8)$, then

$$\inf_{\Psi_n}\max_{0\leq i\leq M}P_{f_i}(\Psi_n\neq i)\geq C(\alpha),$$

where $C(\alpha)>0$ is some finite constant depending only on α.

Hence, combination of Proposition 4.1 and the reduction principle for (4.2) implies the following theorem for verifying lower bounds on the minimax risk in (1.6).

Theorem 4.2. *Let $M\in\mathbb{N}$ and $f_0,\dots,f_M$ be elements of $\mathcal{F}$ with:*

1. *$d(f_i,f_j)\geq 2s_n>0$ for any $0\leq i<j\leq M$ and some sequence $s_n\subset\mathbb{R}_+$;*
2. *$P_{f_i}\ll P_{f_0}$ for any $i=1,\dots,M$, and for some $\alpha\in(0,1/8)$ holds*

$$\frac{1}{M}\sum_{i=1}^{M}\boldsymbol{KL}\left(P_{f_i},P_{f_0}\right)\leq\alpha\max\{\log(M),1\}.$$

Then, for $r_n=s_n/A$ with $A>0$

$$\inf_{T_n}R_w(T_n;\mathcal{F},d,r_n)\geq C(\alpha)w(A),$$

where the infimum is taken over all possible estimators and $C(\alpha)>0$ is a finite constant depending only on α.

4.2 Bivariate boundary fragment model with fixed design

Introduce the function class $\mathcal{M}_{bf}$, which contains all functions $m_\phi:[0,1]^2\to\mathbb{R}$ of the form

$$\begin{aligned}m_\phi(x,y)&=m(x,y)+j_{\tau,\phi}(x,y),\\ j_{\tau,\phi}(x,y)&=j_\tau(x,y)=\tau(x)1_{[0,\phi(x)]}(y),\end{aligned}$$

where $\phi \in C^2[0,1], \tau \in C^2(\mathbb{R}_+)$ and $m \in C^2[0,1]^2$. Note that this is exactly (2.2) with the Assumption 2.2. Define a semi-distance on $\mathcal{M}_{bf}$ by

$$d(m_{\phi_1}, m_{\phi_2}) = \sup_{x \in [0,1]} |\phi_1(x) - \phi_2(x)|, \quad m_{\phi_1}, m_{\phi_2} \in \mathcal{M}_{bf}. \tag{4.3}$$

We can state the following theorem for the minimax rate on $(\mathcal{M}_{bf}, d)$ for model (2.1), which shows that the method in Chapter 2 is minimax optimal up to a logarithmic factor.

Theorem 4.3. *The minimax rate for estimating the jump-location-curve ϕ in model (2.1) on $(\mathcal{M}_{bf}, d)$ is $r_n = 1/n$, where d is defined in (4.3) and $(\epsilon_{i_1,i_2})_{i_1,i_2}$ are centered, independent, normally distributed with standard deviation $\sigma > 0$.*

Proof of Theorem 4.3.

Construction of hypotheses

Let

$$f_0(x,y) = 1_{[0,\theta)}(y), \quad \text{and} \quad f_1(x,y) = 1_{[0,\theta+(2n)^{-1})}(y), \quad x,y \in [0,1],$$

where $\theta \in (0,1)$ is such that $\theta + 1/(2n) \in (0,1)$ for n sufficiently large. Apparently, both hypotheses functions are elements of $\mathcal{M}_{bf}$.

Semi–distance bound

With the definition of the semi distance d in (4.3) it follows that

$$d(f_0, f_1) = \sup_{x \in [0,1]} |\theta - \theta + 1/(2n)| = 1/(2n) =: 2s_n.$$

Kullback–Leibler distance

Note that the distributions P_j of $(Y_{i_1,i_2})_{i_1,i_2}$ with respect to f_j have the following density with respect to the Lebesgue measure on $\mathbb{R}^{n^2}$

$$p_j(y_{11}, \ldots, y_{nn}) = \prod_{i_1,i_2=1}^{n} \phi_\sigma(y_{i_1 i_2} - f_j(\mathbf{x_{i_1,i_2}})), \quad j = 0,1, \tag{4.4}$$

where ϕ_σ denotes the density of a centered normal distribution with standard deviation σ. Further, it holds that

$$(f_0(x,y) - f_1(x,y))^2 = \begin{cases} 1, & y \in [\theta, \theta + 1/(2n)) \\ 0, & \text{else.} \end{cases} \tag{4.5}$$

Now θ can be chosen such that there exists no design point $\mathbf{x_{i_1,i_2}}$ with $(\mathbf{x_{i_1,i_2}})_2 \in [\theta, \theta + 1/(2n))$. With this choice for θ and with (4.4) and (4.5) obtain

$$\begin{aligned}\mathbf{KL}(P_0,P_1) &= \int \log \frac{\mathrm{d}P_0}{\mathrm{d}P_1}\,\mathrm{d}P_0\\ &= \sum_{i_1,i_2=1}^{n} \int \log\left(\frac{\varphi_\sigma(y-f_0(\mathbf{x_{i_1,i_2}}))}{\varphi_\sigma(y-f_1(\mathbf{x_{i_1,i_2}}))}\right)\varphi_\sigma(y-f_0(\mathbf{x_{i_1,i_2}}))\,dy \\ &= \frac{1}{2\sigma^2}\sum_{i_1,i_2=1}^{n}(f_0(\mathbf{x_{i_1,i_2}})-f_1(\mathbf{x_{i_1,i_2}}))^2 = 0,\end{aligned} \tag{4.6}$$

where the last line is due to the explicit formula of the Kullback-Leibler-distance for normal distributions: If $P \sim N(\mu,\sigma^2)$ and $Q \sim N(\tilde{\mu},\tilde{\sigma}^2)$, then

$$\mathbf{KL}(P,Q) = \log\frac{\tilde{\sigma}}{\sigma} + \frac{(\mu-\tilde{\mu})^2+\sigma^2}{2\tilde{\sigma}^2} - \frac{1}{2}.$$

See Belov and Armstrong (2011) for a reference. Theorem 4.2 concludes the proof, since $s_n \cong n^{-1}$. □

4.3 Kink-location with fixed design

In this section we show that the estimate for the kink-location in Chapter 3 is minimax over the function class $\mathcal{M}_s$ as given in Definition 3.1.

Theorem 4.4. *Let $\gamma \in \mathbb{N}$ and $s > 0$ with $s \geq \gamma+1$. Further, let $a,L > 0$ and let $\Theta \subset (0,1)$ be a compact set. Then, for model (3.1) with $\varepsilon_i \sim N(0,\sigma^2)$ it holds that*

$$\liminf_n \inf_{\hat{\theta}} \sup_{m\in\mathcal{M}_s} \mathbb{E}_m\, w\big(n^{(s-\gamma+1)/(2s+1)}\,|\hat{\theta}-\theta_m|\big) \geq C,$$

where $C > 0$ is some constant independent of $\mathcal{M}_s(\gamma,a,\Theta,L)$ and the infimum is taken over all possible estimators $\hat{\theta}$ of θ_m and w is a loss function as in Section 1.1.

Proof of Theorem 4.4.

Construction of the first hypotheses function

Let $\theta_0 \in \Theta$. Define for any $x \in [0,1]$,

$$f_0(x) := \frac{a}{\gamma!}(x-\theta_0)^\gamma 1_{[\theta_0,1]}(x). \tag{4.7}$$

We verify that

$$f_0 \in \mathcal{M}_s(\gamma,a,\Theta,L/2) \subset \mathcal{M}_s(\gamma,a,\Theta,L).$$

To see this, we check the conditions (i) as well as (iia) or (iib) depending on s in Definition 3.1.

Ad (i).

Compute that for $x \in [0,1]$ the $(\gamma-1)$-th derivative of f_0 is $f_0^{(\gamma-1)}(x) = a(x-\theta_0)1_{[\theta_0,1]}(x)$, which is an element of $C^1(\{\theta_0\}^c)$. In addition, for $x \in [0,1]\backslash\{\theta_0\}$ we have $f_0^{(\gamma)}(x) = a1_{[\theta_0,1]}(x)$, such that $[f_0^{(\gamma)}](\theta_0) = a$.

Ad (iia).

If $s = \gamma + 1$, then certainly $f_0^{(\gamma)} \in \mathrm{Lip}(\{\theta_0\}^c, L/2)$.

Ad (iib).

If $s > \gamma + 1$, for $x \in [0,1]\backslash\{\theta_0\}$ we have $f_0^{(\gamma+1)}(x) \equiv 0$, so that $[f_0^{(\gamma+1)}](\theta_0) = 0$. Moreover, it holds trivially that $g_{f_0^{(\gamma)}} \in \mathcal{H}^{s-(\gamma+1)}([0,1], L/2)$.

Construction of the second hypotheses function

For the second hypotheses function define

$$v_0(x) = \begin{cases} 0, & x \in [0,\theta_0), \\ \frac{a}{\gamma!}(x-\theta_0)^\gamma, & x \in [\theta_0,\theta_1], \\ \frac{a}{\gamma!}(\theta_1-\theta_0)^\gamma, & x \in (\theta_1, 1], \end{cases} \tag{4.8}$$

where $\theta_1 = \theta_0 + r_n \in \Theta$ with $r_n = \tilde{C} b_n^{s-\gamma+1}$ and $\tilde{C} > 0$ is a constant chosen below and $b_n = o(1)$ is a real-valued sequence which is as well chosen below. Moreover, let v_n be such that $v_n^{(\gamma)} \in \mathcal{H}^{s-\gamma}([0,1], L/2) \cap C^\infty$. v_n will be constructed explicitly below. The second hypotheses function is set to be

$$f_1 = f_0 - (v_0 - v_n), \tag{4.9}$$

of which we show that it is an element of $\mathcal{M}_s(\gamma, a, \Theta, L)$. Again, we check for f_1 the conditions (i) as well as (iia) or (iib) depending on s in Definition 3.1.

Ad (i).

Note that $f_0 - v_0 = \frac{a}{\gamma!}(x-\theta_1)^\gamma 1_{[\theta_1,1]}$ so that for any $x \in [0,1]$,

$$f_1^{(\gamma-1)}(x) = a(x-\theta_1)1_{[\theta_1,1]}(x) + v_n^{(\gamma-1)}(x)$$

which is an element of $C^1(\{\theta_1\})$ as $v_n^{(\gamma-1)}$ is smooth over $[0,1]$. Further, for any $x \in [0,1]\backslash\{\theta_1\}$,

$$f_1^{(\gamma)}(x) = a1_{[\theta_1,1]}(x) + v_n^{(\gamma)}(x),$$

so that $[f_1^{(\gamma)}](\theta_1) = a$, since $v_n^{(\gamma)}$ is smooth over $[0,1]$.

Ad (iia).

If $s = \gamma + 1$, then $(f_0 - v_0)^{(\gamma)} \in \mathrm{Lip}(\{\theta_1\}^c, L/2)$ and by assumption $v_n^{(\gamma)} \in \mathrm{Lip}([0,1], L/2)$. Hence, the triangle inequality implies $f_1^{(\gamma)} \in \mathrm{Lip}(\{\theta_1\}^c, L/2)$.

Ad (iib).

The assertion can be shown similarly as for f_0 and using the triangle inequality to incorporate v_n.

Construction of v_n

Set

$$v_n(x) = \frac{1}{b_n^{\gamma(s-\gamma+1)-s+1}} \int_{x-b_n}^{x+b_n} v_0(y)\psi((y-x)/b_n)\,\mathrm{d}y, \tag{4.10}$$

where b_n is the same as in (4.8) and which will be explicitly chosen below, and ψ is such that

1. $\mathrm{supp}(\psi) = (-1,1)$, $\psi(\pm 1) = 0$ and $\int \psi = 1$.
2. ψ is infinitely often continuous differentiable inside its support.
3. There exists a finite constant $C_{R,\psi} > 0$ such that $||\psi||_\infty < C_{R,\psi}$.
4. $\psi \in \mathcal{H}^s([-1,1], C_{L,\psi})$ for some finite constant $C_{L,\psi} > 0$.

Note that $\psi^{(\gamma)}(x) = \tilde{C}_\psi \exp(-1/(1-x^2))1_{[-1,1]}$ with $\tilde{C}_\psi$ such that $\int \psi^{(\gamma)} = 1$ satisfies all the assumption 1.–4. Thus, v_n is a smooth version of v_0.

With the choice (4.10) for v_n we have that $v_n^{(\gamma)} \in \mathcal{H}^{s-\gamma}([-1,1], L/2) \cap C^\infty$ Indeed, v_n is infinitely often differentiable with

$$v_n^{(j)}(x) = (-1)^j \frac{1}{b_n^{\gamma(s-\gamma+1)-s+1+j}} \int_{x-b_n}^{x+b_n} v_0(t)\psi^{(j)}((t-x)/b_n)\,\mathrm{d}t, \quad j \in \mathbb{N}, x \in [0,1].$$

Note that by definition of v_0 in (4.8)

$$\sup_{x\in[0,1]} v_0(x) = \frac{a}{\gamma!}(\theta_1 - \theta_0)^\gamma = \frac{a}{\gamma!} r_n^\gamma = \frac{a\tilde{C}^\gamma}{\gamma!} b_n^{\gamma(s-\gamma+1)}. \tag{4.11}$$

Thus, with (4.11) and the properties of ψ, one has for any $x, y \in [0,1]$

$$\begin{aligned}
&|v_n^{(\lfloor s\rfloor)}(x) - v_n^{(\lfloor s\rfloor)}(y)| \\
&\quad = \frac{1}{b_n^{\gamma(s-\gamma+1)-s+\lfloor s\rfloor+1}} \Big| \int_{B_{b_n}(x)\cup B_{b_n}(y)} v_0(t)\big(\psi^{(\lfloor s\rfloor)}((t-x)/b_n) - \psi^{(\lfloor s\rfloor)}((t-y)/b_n)\big)\,\mathrm{d}t \Big| \\
&\quad \le \frac{C_{L,\psi}|x-y|^{s-\lfloor s\rfloor}}{b_n^{\gamma(s-\gamma+1)+1}} \int_{B_{b_n}(x)\cup B_{b_n}(y)} |v_0(t)|\,\mathrm{d}t \\
&\quad \le \frac{4C_{L,\psi}\,a\tilde{C}^\gamma}{\gamma!}|x-y|^{s-\lfloor s\rfloor},
\end{aligned}$$

where $B_\rho(x) := [x-\rho, x+\rho]$ for $x \in \mathbb{R}$ and $\rho > 0$. Consequently, we choose $\tilde{C}$ such that ${}^{4C_{L,\psi}\,a\tilde{C}^\gamma}/_{\gamma!} \le L/2$, which implies $v_n^{(\gamma)} \in \mathcal{H}^{s-\gamma}([-1,1], L/2) \cap C^\infty$.

Likewise, with (4.11) one easily derives that

$$||v_n||_\infty \le \frac{C_{R,\psi}\,a\tilde{C}^\gamma}{\gamma!} b_n^s. \tag{4.12}$$

Semi-distance bound

We define a semi-distance d on $\mathcal{M}_s(\gamma, a, \Theta, L)$ by

$$d(f_1, f_2) = |\theta_{f_1} - \theta_{f_2}|, \quad f_1, f_2 \in \mathcal{M}_s(\gamma, a, \Theta, L),$$

where for $i = 1,2$ the unique γ-kink of f_i is θ_{f_i}. As a result, by choice of f_0 and f_1 in (4.7) and (4.9) obtain that

$$d(f_0, f_1) = |\theta_0 - \theta_1| = r_n =: 2s_n.$$

Kullback-Leibler-distance

By (4.11) and (4.12) for any $x \in [0,1]$ it holds that

$$(v_n(x) - v_0(x))^2 = \leq 2\Big(\frac{C_{R,\psi}\, a\, \tilde{C}^\gamma}{\gamma!}\Big)^2 b_n^{2s} + 2\Big(\frac{a\,\tilde{C}^\gamma}{\gamma!}\Big)^2 b_n^{2\gamma(s-\gamma+1)}.$$

Hence, in view of (4.7) and (4.9), there exists a finite constant $\tilde{C}_{\psi,a,\gamma} > 0$ depending only on ψ, a and γ such that

$$\sup_{x\in[0,1]} (f_0(x) - f_1(x))^2 \leq \tilde{C}_{\psi,a,\gamma}\, b_n^{2s}, \tag{4.13}$$

as $s \geq \gamma + 1 \geq 2$. Further, for $x \in [0, \theta_0 - b_n) \cup (\theta_1 + b_n, 1]$ it holds that $v_n - v_0 \equiv 0$. To see this, note that if $x \in [0, \theta_0 - b_n)$ then $v_0(x) = 0$ and

$$v_n(x) = b_n^{-1} \int_{x-b_n}^{x+b_n} v_0(y)\psi((y-x)/b_n)\,\mathrm{d}y = 0.$$

Otherwise, if $x \in (\theta_1 + b_n, 1]$ then $v_0(x) = \frac{a}{\gamma!}(\theta_1 - \theta_0)^\gamma$ and

$$v_n(x) = \frac{a}{\gamma!}(\theta_1 - \theta_0)^\gamma\, b_n^{-1} \int_{x-b_n}^{x+b_n} \psi((y-x)/b_n)\,\mathrm{d}y = \frac{a}{\gamma!}(\theta_1 - \theta_0)^\gamma = v_0(x).$$

Thus,

$$\sum_{i=1}^n (f_0(x_i) - f_1(x_i))^2 \leq \sup_{x\in[0,1]} (f_0(x) - f_1(x))^2 \cdot 2n(r_n + 2b_n), \tag{4.14}$$

since only for the design points $x_i \in [\theta_0 - b_n, \theta_1 + b_n]$ the summands are not equal to zero, which are less than $2n(\theta_1 + b_n - (\theta_0 - b_n)) = 2n(r_n + 2b_n)$ summands due to the equidistant design. Thus, (4.13) and (4.14) imply for n large enough, and possibly making the constant $\tilde{C}_{\psi,a,\gamma}$ larger, that

$$\sum_{i=1}^n (f_0(x_i) - f_1(x_i))^2 < \tilde{C}_{\psi,a,\gamma}\, b_n^{2s+1}\, n, \tag{4.15}$$

since r_n is asymptotically negligible compared to b_n.

We turn to the bound on the Kullback-Leibler-distance. For $j = 0, 1$ the distributions P_j of $Y_1, \ldots, Y_n$ with respect to f_j have a density with respect to the Lebesgue measure on $\mathbb{R}^n$ with a similar shape as in (4.4). Hence, the Kullback-Leibler-divergence can be computed similarly as in (4.6) to $\mathbf{KL}(P_0, P_1) = \frac{1}{2\sigma^2}\sum_{i=1}^n (f_0(x_i) - f_1(x_i))^2$. By means of (4.15) it holds for sufficiently large n that

$$\mathbf{KL}(P_0, P_1) \leq \frac{\tilde{C}_{\psi,a,\gamma}}{2\sigma^2}\, b_n^{2s+1}\, n.$$

Now, we choose b_n such that the right-hand side of the latter display is smaller than some fixed constant $\alpha \in (0, 1/8)$, which is the case by choosing $b_n \cong n^{-1/(2s+1)}$. Theorem 4.2 implies that the minimax lower bound over the function class $\mathcal{M}_s$ is of order $s_n \cong r_n \cong b_n^{s-\gamma+1} \cong n^{-(s-\gamma+1)/(2s+1)}$ as claimed.

□

APPENDIX A

Gaussian and sub-Gaussian processes

A.1 Properties of Gaussian processes

We consider in this section the design of the model in (2.1), i.e. $(\mathbf{x_{i_1,i_2}})_{i_1,i_2=1,\dots,n}$ form a deterministic, regular rectangular grid in $[0,1]^2$. For $x \in [0,1]$ set $\mathbf{p}(x) = (x, \phi(x))^T$ for some function $\phi \in C^2[0,1]$ with image in $(0,1)$ and for sake of brevity we write

$$\mathbf{p}_{w,h}(x) = \mathbf{p}(x) + hw\mathbf{e}_2 = (x, \phi(x) + hw)^T,$$

where $h < 1/2$ is a bandwidth parameter tending to zero for $n \to \infty$. Further, we set $\Theta_n = \bigcup_{x \in I} \{x\} \times \tilde{\Theta}_{n,x}$, where $I \subset (0,1)$ is a compact set and

$$\tilde{\Theta}_{n,x} = \left\{ w \in \mathbb{R} : \phi(x) + wh \in \left[h, 1-h\right] \right\} \times [-\pi/2, \pi/2].$$

Let $f : \mathbb{R}^2 \to \mathbb{R}^{2\times 2}$ be a function with $f \in C^1(\mathbb{R}^2)$ having compact support in $[-1,1]^2$ and bounded marginals. Moreover, let $g_1 : \mathbb{R}^2 \to \mathbb{R}$ be a function in $C^1(\mathbb{R}^2)$ which does not depend on the bandwidth h and $g_2, g_3 : \mathbb{R}^2 \to \mathbb{R}^2$ be functions which have the following shape: For $i = 2,3$ let either $g_i(\mathbf{z}) = \mathbf{C}_{m,g_i}(\mathbf{p}_{w,h}(x) - \mathbf{z})$, for some constant matrix $\mathbf{C}_{m,g_i} \in \mathbb{R}^{2\times 2}$, or $g_i(\mathbf{z}) \equiv \mathbf{C}_{g_i}$, for some constant vector $\mathbf{C}_{g_i} \in \mathbb{R}^2$. Both $\mathbf{C}_{m,g_i}$ and $\mathbf{C}_{g_i}$ can depend on x or ψ. In addition, let $r_2, r_3 \in \{0,1\}$ be such that $r_i = 0$ if and only if $g_i(\mathbf{z}) \equiv \mathbf{C}_{g_i}$, and $r_i = 1$ if and only if $g_i(\mathbf{z}) = \mathbf{C}_{m,g_i}(\mathbf{p}_{w,h}(x) - \mathbf{z})$, for $i = 2,3$ respectively. Furthermore, for $i = 2,3$ let $\tilde{g}_i : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

$$\tilde{g}_i(\mathbf{z}) = \begin{cases} g_i(\mathbf{z}), & \text{if } g_i(\mathbf{z}) \equiv \mathbf{C}_{g_i}, \text{ or} \\ (-1)^{r_i} \mathbf{C}_{m,g_i}(\mathbf{D}_\psi \mathbf{z}), & \text{else,} \end{cases}$$

where $\mathbf{D}_\psi$ denotes the rotation matrix (1.13). Note that these functions do not dependent on h and

$$g_i(\mathbf{p}_{w,h}(x) - h\mathbf{D}_\psi \mathbf{z}) = h^{r_i} \tilde{g}_i(\mathbf{z}), \quad i = 2,3. \tag{A.1}$$

Let $r_1 = 1 + r_2 + r_3$ and let $(\boldsymbol{\varepsilon}_{i_1,i_2})_{i_1,i_2}$ be square-integrable, centered, independent and identically distributed random variables with standard deviation $\sigma > 0$. For $(x, w, \psi)^T \in \Theta_n$ set

$$Z_{n,0}(x,w,\psi) = \frac{1}{nh^{r_1}\sigma} \sum_{i_1,i_2=1}^{n} \boldsymbol{\varepsilon}_{i_1,i_2}\, g_1(\mathbf{x_{i_1,i_2}}) \left\langle f\left(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x) - \mathbf{x_{i_1,i_2}})\right) g_2(\mathbf{x_{i_1,i_2}}), g_3(\mathbf{x_{i_1,i_2}}) \right\rangle \tag{A.2}$$

and

$$Z_{n,G}(x,w,\psi) := \frac{1}{h^{r_2+r_3}} \int_{\mathbb{R}^2} g_1(h\mathbf{z}) \big\langle f\big(\mathbf{D}_{-\psi}(\mathbf{p}(x)/h + w\mathbf{e}_2 - \mathbf{z})\big)\, g_2(h\mathbf{z}), g_3(h\mathbf{z}) \big\rangle \, \mathrm{d}W(\mathbf{z}), \tag{A.3}$$

where W is a Wiener sheet on $\mathbb{R}^2$.

Gaussian approximation

Lemma A.1. *Suppose the errors $\boldsymbol{\varepsilon}_{i_1,i_2}$ are such that $\mathbb{E}|\varepsilon_{1,1}|^5 < \infty$. In addition, if $h = n^{-\eta}$, where $\eta \in (0,1/2)$, then on an appropriate probability space there exists a Wiener Sheet on $\mathbb{R}^2$ such that for sufficiently large n holds*

$$\sup_{(x,w,\psi)^T \in \Theta_n} |Z_{n,0}(x,w,\psi) - Z_{n,G}(x,w,\psi)| = O_P\Big(\frac{\sqrt{\log(n)}}{n^{1/2}h}\Big).$$

In particular,

$$\sup_{(x,w,\psi)^T \in \Theta_n} |Z_{n,0}(x,w,\psi)| - \sup_{(x,w,\psi)^T \in \Theta_n} |Z_{n,G}(x,w,\psi)| = O_P\Big(\frac{\sqrt{\log(n)}}{n^{1/2}h}\Big).$$

The proof is provided in Section A.1.1.

It easily follows a more simpler result for the one-dimensional case, that is considering an equidistant design $(x_i)_{i=1,\dots,n}$ on $[0,1]$ and the noise variables $(\varepsilon_i)_{i=1,\dots,n}$ have similar distributional properties as $(\boldsymbol{\varepsilon}_{i_1,i_2})_{i_1,i_2}$ before.

Lemma A.2. *Let $(\varepsilon_i)_{i=1,\dots,n}$ be square-integrable, centered, independent and identically distributed random variables with standard deviation $\sigma > 0$ and moreover $\mathbb{E}|\varepsilon_1|^5 < \infty$. In addition, suppose that $h = n^{-\eta}$, where $\eta \in (0,1/2)$. Further, let K be two-times continuous differentiable with support in $[-1,1]$, then on a suitable probability space for n sufficiently large,*

$$\sup_{x\in[h,1-h]} |Z_n(x) - Z_G(x)| = O_P\Big(\sqrt{\frac{\log(n)}{nh}}\Big),$$

where

$$Z_n(x) = \frac{1}{\sqrt{nh}\sigma} \sum_{i=1}^{n} \varepsilon_i K(h^{-1}(x-x_i)), \quad \textit{and} \quad Z_G(x) = h^{-1/2} \int_{\mathbb{R}} K(h^{-1}(x-t))\, dW(t)$$

and W is a Brownian motion on $\mathbb{R}$.

Moments and quantiles of Gaussian processes

Lemma A.3. *Suppose the errors $\boldsymbol{\varepsilon}_{i_1,i_2}$ are such that $\mathbb{E}|\varepsilon_{1,1}|^5 < \infty$. In addition, if $h = n^{-\eta}$, where $\eta \in (0,1/2)$, then the following statements are true.*

1. *For any $\zeta > C_0$, where $C_0 > 0$ is a finite constant uniform for x, w and ψ, it holds*

$$P\big(\sup_{(x,w,\psi)\in\Theta_n} |Z_{n,G}(x,w,\psi)| > \zeta\big) \le C_1 \frac{\lambda_3(\Theta_n)\zeta}{h^3} \exp(-C_2\zeta^2)$$

where C_1, C_2 are finite absolute constants uniform for x, w and ψ.

2. *There exists a finite constant $C_3 > 0$ uniform for x, w and ψ such that for sufficiently small h (sufficiently large n) holds*

$$\mathbb{E} \sup_{(x,w,\psi)\in\Theta_n} |Z_{n,G}(x,w,\psi)| \leq C_3 \sqrt{\log(n)}.$$

3. *There exists a finite constant $C_4 > 0$ uniform for x, w and ψ such that for any $\delta \in (0,1)$*

$$\mathbb{E} \sup_{\theta_1,\theta_2\in\Theta_n \,:\, ||\theta_1-\theta_2||\leq\delta} |Z_{n,G}(\theta_1) - Z_{n,G}(\theta_2)| \leq C_4 \, \delta \, h^{-1} \sqrt{\log(n)}.$$

4. *For any $\alpha \in (0,1)$*

$$q_\alpha\Big(\sup_{(x,w,\psi)\in\Theta_n} |Z_{n,G}(x,0,\psi(x))| \Big) \cong \sqrt{\log(n)}.$$

The proof is deferred to Section A.1.2.

A.1.1 Proof of Lemma A.1

The following result can be found in Proksch et al. (2015).

Lemma A.4 (Integration by parts for Wiener sheet integrals)**.** *Let $[a_1,b_1]\times[a_2,b_2] \subset \mathbb{R}^2_+$. For a continuously differentiable function $f : \mathbb{R}^2 \to \mathbb{R}$ and a standard Wiener sheet on $[0,\infty)^2$ it holds that*

$$\begin{aligned}
\int_{[a_1,b_1]\times[a_2,b_2]} f(z_1,z_2)\,\mathrm{d}W(z_1,z_2) = &\int_{[a_1,b_1]\times[a_2,b_2]} W(z_1,z_2) f^{(1,1)}(z_1,z_2)\,\mathrm{d}z_1\mathrm{d}z_2 \\
&- \int_{[a_1,b_1]} W(z_1,b_2) f^{(1,0)}(z_1,b_2) + W(z_1,a_2) f^{(1,0)}(z_1,a_2)\,\mathrm{d}z_1 \\
&- \int_{[a_2,b_2]} W(b_1,z_2) f^{(0,1)}(b_1,z_2)\,\mathrm{d}z_2 + W(a_1,z_2) f^{(0,1)}(a_1,z_2)\,\mathrm{d}z_2 \\
&+ W(b_1,b_2) f(b_1,b_2) - W(a_1,b_2) f(a_1,b_2) - W(b_1,a_2) f(b_1,a_2) \\
&+ W(a_1,a_2) f(a_1,a_2).
\end{aligned}$$

*Proof of Lemma A.*1. For sake of brevity write for $\mathbf{z} \in \mathbb{R}^2$,

$$F((h^{-1}\mathbf{D}_{-\psi}\left(\mathbf{p}_{w,h}(x) - \mathbf{z}\right); w, \psi, h) := g_1(\mathbf{z}) \left\langle f\left(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x) - \mathbf{z})\right) g_2(\mathbf{z}), g_3(\mathbf{z}) \right\rangle.$$

Note that F is a real-valued, bounded and continuous differentiable function with the same compact support as f, that is $[-1,1]^2$. In the following we will suppress the dependency of F on w, ψ and h in the notation. Furthermore,

$$F(\mathbf{z}) = g_1(\mathbf{p}_{w,h}(x) + h\mathbf{D}_\psi \mathbf{z}) \left\langle f(\mathbf{z}) g_2(\mathbf{p}_{w,h}(x) + h\mathbf{D}_\psi \mathbf{z}), g_3(\mathbf{p}_{w,h}(x) + h\mathbf{D}_\psi \mathbf{z}) \right\rangle,$$

so that

$$||F||_\infty = O(h^{r_2+r_3}), \quad \text{and} \quad ||F^{(1,1)}||_\infty = O(h^{r_2+r_3}), \tag{A.4}$$

uniformly for x, w and ψ, due to the shapes of g_i, $i = 1,2,3$. This uniformity comes from the fact that x and ψ are element of compact sets and w is such that $\mathbf{p}_{w,h}(x)$ is element of a compact subset.

Step 1: Change summation order

Define $S_{i_1,i_2} = \sum_{l=1}^{i_1}\sum_{k=1}^{i_2} \varepsilon_{l,k}$ for $(i_1,i_2) \in \{0,1,\ldots,n\}^2$ and set $S_{i_1,0} = S_{0,i_2} = 0$ for all $(i_1,i_2) \in \{0,1,\ldots,n\}^2$. Thus,

$$\boldsymbol{\varepsilon}_{i_1,i_2} = S_{i_1,i_2} - S_{i_1-1,i_2} - S_{i_1,i_2-1} + S_{i_1-1,i_2-1}$$

and we get

$$Z_{n,0}(x,w,\psi) = \frac{1}{nh^{r_1}\sigma}\sum_{i_1,i_2=1}^{n} F(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x)-\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}))(S_{i_1,i_2} - S_{i_1-1,i_2} - S_{i_1,i_2-1} + S_{i_1-1,i_2-1}).$$

Rewrite the process $Z_{0,n,h}$ as

$$\begin{aligned} Z_{n,0}(x,w,\psi) = \frac{1}{nh^{r_1}\sigma}\sum_{i_1,i_2=1}^{n-1} \Big(& F(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x)-\mathbf{x}_{i_1+1,i_2+1})) - F(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x)-\mathbf{x}_{i_1,i_2+1})) \\ & - F(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x)-\mathbf{x}_{i_1+1,i_2})) + F(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x)-\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}))\Big) S_{i_1,i_2} + \frac{R_{0,n}(x,w,\psi)}{nh^{r_1}\sigma}, \end{aligned}$$

where

$$\begin{aligned} R_{0,n}(x,w,\psi) = & -\sum_{i_1=1}^{n-1}\Big(F(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x)-\mathbf{x}_{i_1+1,n})) - F(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x)-\mathbf{x}_{i_1,n}))\Big) S_{i_1,n} \\ & -\sum_{i_2=1}^{n-1}\Big(F(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x)-\mathbf{x}_{n,i_2+1})) - F(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x)-\mathbf{x}_{n,i_2}))\Big) S_{n,i_2} \\ & + F(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x)-x_{n,n}))S_{n,n}. \end{aligned}$$

Because $(x,w,\psi)^T \in \Theta_n$ it holds that $\phi(x)+hw \in [h,1-h]$. Additionally, all occurring design points in $R_{0,n}(x,w,\psi)$ are (nearly) edge-points such that for sufficiently large n all terms in $R_{0,n}(x,w,\psi)$ vanish as the arguments in the inverse image of F leave the compact support of F. This vanishing property holds uniformly over Θ_n for any n sufficiently large. Thus, for sufficiently large n

$$\begin{aligned} Z_{n,0}(x,w,\psi) = \frac{1}{nh^{r_1}\sigma}\sum_{i_1,i_2=1}^{n-1} \Big(& F(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x)-\mathbf{x}_{i_1+1,i_2+1})) - F(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x)-\mathbf{x}_{i_1,i_2+1})) \\ & - F(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x)-\mathbf{x}_{i_1+1,i_2})) + F(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x)-\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}))\Big) S_{i_1,i_2}. \end{aligned}$$

Step 2: Approximation by a Wiener Sheet

Introduce the process

$$\begin{aligned} Z_{n,1}(x,w,\psi) = \frac{1}{nh^{r_1}}\sum_{i_1,i_2=1}^{n-1} \Big(& F(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x)-\mathbf{x}_{i_1+1,i_2+1})) - F(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x)-\mathbf{x}_{i_1,i_2+1})) \\ & - F(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x)-\mathbf{x}_{i_1+1,i_2})) + F(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x)-\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}))\Big) W(i_1,i_2), \end{aligned}$$

where W is a Wiener Sheet with

$$\sup_{1\le i_1,i_2\le n} |S_{i_1,i_2}\sigma^{-1} - W(i_1,i_2)| = O_P\left(n^{\frac{2-\delta}{2}}\sqrt{\log(n)}\right). \tag{A.5}$$

Such a Wiener Sheet exists, see Theorem 1 in Rio (1993), provided there exists some $\delta \in (0,1]$ such that $\mathbb{E}|\varepsilon_{1,1}|^k < \infty$ for $k > 4/(2-\delta)$. Note that the preconditions of this lemma allow to choose $\delta = 1$. Let

$$A_{i_1,i_2} = [x_{i_1}, x_{i_1+1})\times[x_{i_2}, x_{i_2+1}), \quad i_1,i_2 \in \{1,\dots,n-1\},$$

where the right boundary is included if $i_1 = n-1$ or $i_2 = n-1$. This yields

$$\begin{aligned}
&|Z_{n,0}(x,w,\psi) - Z_{n,1}(x,w,\psi)| \\
&\le \sup_{1\le i_1,i_2\le n} \frac{|S_{i_1,i_2}\sigma^{-1} - W(i_1,i_2)|}{nh^{r_1}} \sum_{i_1,i_2=1}^{n-1} |F(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x)-\mathbf{x}_{i_1+1,i_2+1})) - F(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x)-\mathbf{x}_{i_1,i_2+1})) \\
&\qquad - F(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x)-\mathbf{x}_{i_1+1,i_2})) + F(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x)-\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}))| \\
&\le \sup_{1\le i_1,i_2\le n} \frac{|S_{i_1,i_2}\sigma^{-1} - W(i_1,i_2)|}{nh^{r_1}} \sum_{i_1,i_2=1}^{n-1} \int_{A_{i_1,i_2}} h^{-2}|F^{(1,1)}(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x)-\mathbf{z}))|\,\mathrm{d}\mathbf{z} \\
&\le \sup_{1\le i_1,i_2\le n} \frac{|S_{i_1,i_2}\sigma^{-1} - W(i_1,i_2)|}{nh^{r_1}} \int_{[-1,1]^2} |F^{(1,1)}(\mathbf{z})|\,\mathrm{d}\mathbf{z} = O_P\left(n^{\frac{-1}{2}} h^{r_2+r_3-r_1}\sqrt{\log(n)}\right) \\
&= O_P\left(n^{\frac{-1}{2}} h^{-1}\sqrt{\log(n)}\right),
\end{aligned}$$

uniformly for x, w and ψ, where we used for the second last equality (A.4) and (A.5), while the last equality is due to the choice of r_1 (see the text above (A.2)).

Step 3: Continuous approximation

Next, introduce the process

$$Z_{n,2}(x,w,\psi) = \frac{1}{h^{r_1}} \int\limits_{[0,1]^2} F(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x)-\mathbf{z}))\,\mathrm{d}W(\mathbf{z}).$$

Hence, by integration by parts for the Wiener sheet (Lemma A.4),

$$Z_{n,2}(x,w,\psi) = \frac{1}{h^{r_1}} \int\limits_{[0,1]^2} h^{-2}F^{(1,1)}(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x)-z))W(\mathbf{z})\,\mathrm{d}\mathbf{z} + \frac{1}{h^{r_1}} R_{n,1}(x,w,\psi),$$

where

$$\begin{aligned}
R_{n,1}(x,w,\psi) = &-\int\limits_{[0,1]} h^{-1}W(z_1,1)F^{(1,0)}(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x)-(z_1,1)^T))\,\mathrm{d}z_1 \\
&-\int\limits_{[0,1]} h^{-1}W(1,z_2)F^{(0,1)}(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x)-(1,z_2)^T))\,\mathrm{d}z_2 \\
&+W(1,1))F(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x)-(1,1)^T)).
\end{aligned}$$

For sufficiently large n one has that $R_{n,1}(x,w,\psi) \equiv 0$ as the support of the partial derivatives of F and F itself will be exceeded resp. deceeded. This holds uniformly over Θ_n for any n large enough.

Furthermore, obtain by the scaling properties of the Wiener sheet

$$\begin{aligned}Z_{n,1}(x,w,\psi) &= \frac{1}{nh^{r_1}} \sum_{i_1,i_2=1}^{n-1} W(i_1,i_2) \int\limits_{A_{i_1,i_2}} h^{-2}(F^{(1,1)} h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x)-z))\,\mathrm{d}\mathbf{z} \\ &\stackrel{d}{=} \frac{1}{h^{r_1}} \sum_{i_1,i_2=1}^{n-1} W(i_1/n,i_2/n) \int\limits_{A_{i_1,i_2}} h^{-2}(F^{(1,1)} h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x)-z))\,\mathrm{d}\mathbf{z}\end{aligned}$$

and hence for large enough n

$$\begin{aligned}&|Z_{n,1}(x,w,\psi) - Z_{n,2}(x,w,\psi)| \\ &\stackrel{d}{=} \frac{1}{h^{r_1}} \Big| \sum_{i_1,i_2=1}^{n-1} h^{-2} \int\limits_{A_{i_1,i_2}} F^{(1,1)}(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x)-\mathbf{z}))[W(\mathbf{z}) - W(i_1/n,i_2/n)]\,\mathrm{d}\mathbf{z}\Big|.\end{aligned}$$

A modulus of continuity for the Wiener sheet is given in Theorem 1 in Khoshnevisan (2002), that is for any $\delta \in (0,1)$

$$\sup_{||\mathbf{z}_1-\mathbf{z}_2||_\infty<\delta} |W(\mathbf{z}_1) - W(\mathbf{z}_2)| = O_P\left(\sqrt{\log(1/\delta)\delta}\right). \tag{A.6}$$

In addition, for a finite constant $C > 0$ which is uniform in x, w and ψ we have by substitution and (A.4) that

$$\sum_{i_1,i_2=1}^{n-1} \int\limits_{A_{i_1,i_2}} h^{-2}\, |F^{(1,1)}(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x)-\mathbf{z}))|\mathrm{d}\mathbf{z} \le \int_{[-1,1]^2} |F^{(1,1)}(\mathbf{z})|\mathrm{d}\mathbf{z} < Ch^{r_2+r_3}$$

and therefore by incorporating the equidistant design in (A.6) one has

$$|Z_{n,1}(x,w,\psi) - Z_{n,2}(x,w,\psi)| = O_P\left(\sqrt{\log(n)}n^{-1/2}h^{r_2+r_3-r_1}\right) = O_P\left(\sqrt{\log(n)}n^{-1/2}h^{-1}\right),$$

where the O-term is uniform in x, w and ψ.

Step 4: Extension of the support

Approximate $Z_{n,2}(x,w,\psi)$ by $Z_{n,G}(x,w,\psi)$ as in (A.3). For this purpose, obtain by a substitution and the scaling properties of the Wiener sheet that

$$Z_{n,2}(x,w,\psi) \stackrel{d}{=} \frac{1}{h^{r_1-1}} \int\limits_{[0,\frac{1}{h}]^2} F(\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x)/h - \mathbf{z}))\,\mathrm{d}W(\mathbf{z})$$

and with this

$$Z_{n,G}(x,w,\psi) - Z_{n,2}(x,w,\psi) \stackrel{d}{=} \frac{1}{h^{r_1-1}} \int\limits_{A_n} F(\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x)/h - z))\,\mathrm{d}W(\mathbf{z}) =: R_{n,2}(x,w,\psi),$$

where $A_n = \mathbb{R}^2 \setminus [0,h^{-1}]^2$. For $\mathbf{z} = (z_1,z_2)^T \in A_n$ one has that $z_2 < 0$ or $z_2 > h^{-1}$. Moreover, as $(x,w,\psi)^T \in \Theta_n$, it holds that $\phi(x)/h + w \in [1,h^{-1}-1]$ Thus, $\phi(x)/h + w - z_2$ leaves the support of the function F for large enough n. This yields $R_{n,2}(x,w,\psi) \equiv 0$ for sufficiently large n uniformly over Θ_n.

Step 5: Conclusion

Summarizing all approximations steps above, we have for sufficiently large n

$$\begin{aligned}
&Z_{n,0}(x,w,\psi) - Z_{n,G}(x,w,\psi)\\
&\quad \leq |Z_{n,0}(x,w,\psi) - Z_{n,1}(x,w,\psi)| + Z_{n,1}(x,w,\psi) - Z_{n,G}(x,w,\psi)\\
&\quad \leq O_P\Big(\frac{\sqrt{\log(n)}}{n^{1/2}h}\Big) + |Z_{n,1}(x,w,\psi) - Z_{n,2}(x,w,\psi)| + Z_{n,2}(x,w,\psi) - Z_{n,G}(x,w,\psi)\\
&\quad \leq O_P\Big(\frac{\sqrt{\log(n)}}{n^{1/2}h}\Big) + O_P\Big(\frac{\sqrt{\log(n)}}{\sqrt{n}h}\Big) + R_{n,2}(x,w,\psi)\\
&\quad = O_P\Big(\frac{\sqrt{\log(n)}}{n^{1/2}h}\Big),
\end{aligned}$$

uniformly for x,w and ψ. Additionally,

$$F(\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x)/h - \mathbf{z})) = g_1(h\mathbf{z})\Big\langle f(\mathbf{D}_{-\psi}(\mathbf{p}(x)/h + w\mathbf{e}_2 - \mathbf{z}))g_2(h\mathbf{z}), g_3(h\mathbf{z})\Big\rangle,$$

which concludes the first assertion of the lemma. The second assertion follows by the triangle inequality. $\square$

A.1.2 Proof of Lemma A.3

Lemma A.5. *For any $\varepsilon \in (0,1)$, $\rho < 1$ and $\eta > 0$ there exists a finite constant $C > 0$ depending only on ρ and η such that*

$$\int_0^{\varepsilon} x^{-\rho} \log(x^{-1})^{\eta}\, \mathrm{d}x \leq C|\log(\varepsilon^{-1})|^{\eta}\varepsilon^{1-\rho}.$$

Proof. Substituting $t = \log(x^{-1})^{\eta}$, one has $x = \exp(-t^{1/\eta})$ and $\mathrm{d}x = -\frac{x}{\eta}\frac{1}{\log(x^{-1})^{\eta-1}}\,\mathrm{d}t$. Thus,

$$\int_0^{\varepsilon} x^{-\rho}\log(x^{-1})^{\eta}\,\mathrm{d}x = \int_{\log(\varepsilon^{-1})^{\eta}}^{\infty} \frac{1}{\eta} t^{1/\eta}\exp((\rho-1)t^{1/\eta})\,\mathrm{d}t.$$

Substituting $z = t^{1/\eta}$, one has $t = z^{\eta}$ and $\mathrm{d}t = \eta \cdot t^{-\frac{1-\eta}{\eta}}\,\mathrm{d}z$. Therefore,

$$\begin{aligned}
\int_0^{\varepsilon} x^{-\rho}\log(x^{-1})^{\eta}\,\mathrm{d}x &= \int_{\log(\varepsilon^{-1})}^{\infty} z^{\eta}\exp((\rho-1)z)\,\mathrm{d}z\\
&= \Big[\frac{z^{\eta}\exp((\rho-1)z)}{\rho-1}\Big|_{\log(\varepsilon^{-1})}^{\infty} - \int_{\log(\varepsilon^{-1})}^{\infty} \eta\frac{z^{\eta-1}\exp((\rho-1)z)}{(\rho-1)}\,\mathrm{d}z\\
&= -\frac{\log(\varepsilon^{-1})^{\eta}}{(\rho-1)}\varepsilon^{1-\rho} - \frac{\eta}{(\rho-1)}\int_{\log(\varepsilon^{-1})}^{\infty} z^{\eta-1}\exp((\rho-1)z)\,\mathrm{d}z.
\end{aligned}$$

Emplying iteratively integration by parts one obtains that the integral $\int_{\log(\varepsilon^{-1})}^{\infty} z^{\eta}\exp((\rho-1)z)dz$ is equal to

$$\sum_{k=1}^{n}(-1)^{k-1}\frac{\prod_{j=2}^{n}(\eta-j+2)}{(\rho-1)^k}\log(\varepsilon^{-1})^{\eta-k+1}\varepsilon^{1-\rho}+(-1)^n\frac{\eta(\eta-1)\dots(\eta-n+1)}{(\rho-1)^n}\int\limits_{\log(\varepsilon^{-1})}^{\infty} z^{\eta-n}\exp((\rho-1)z)dz,$$

for any $n\in\mathbb{N}$. Note that all terms are bounded by $|\log(\varepsilon^{-1})|^{\eta}\varepsilon^{1-\rho}$ up to a multiplicative constant depending only on ρ and η. This concludes the proof. □

Lemma A.6. *It holds that*

$$q_{\alpha}\Big(\sup_{(x,w,\psi)\in\Theta_n}|Z_{n,G}(x,w,\psi)|\Big)\cong\mathbb{E}\Big(\sup_{(x,w,\psi)\in\Theta_n}|Z_{n,G}(x,w,\psi)|\Big).$$

Proof. For sake of brevity write $\mathbf{M}=\sup_{(x,w,\psi)\in\Theta_n}|Z_{n,G}(x,w,\psi)|$. Without loss of generality assume that $\mathbb{E}(\mathbf{M}^2)\le 1$. Use Lemma B.1 in Chernozhukov et al. (2014) to obtain for any $\alpha\in(0,1)$

$$q_{\alpha}(\mathbf{M})\le C\,\mathbb{E}(\mathbf{M}),$$

for some constant $C>0$. Let $C_1=\sqrt{2|\log(1/\alpha)|}$, then Borell's inequality (see Proposition A.2.1 in van der Vaart and Wellner (1996)) implies

$$P(\mathbf{M}\le\mathbb{E}(\mathbf{M})-C_1)\le P(|\mathbf{M}-\mathbb{E}(\mathbf{M})|\ge C_1)\le\exp(-C_1^2/2)=\alpha.$$

This yields $q_{\alpha}(\mathbf{M})\ge c\,\mathbb{E}(\mathbf{M})$ for some constant $c>0$. □

Proof of Lemma A.3. First part: Maximal inequality

We intend to use Proposition A.2.7 in van der Vaart and Wellner (1996). For this purpose, define the following semi-metric on Θ_n

$$\rho^2\big((x_1,w_1,\psi_1)^T,(x_2,w_2,\psi_2)^T\big)=\mathbb{E}|Z_{n,G}(x_1,w_1,\psi_1)-Z_{n,G}(x_2,w_2,\psi_2)|^2.$$

Note that this semi-metric depends on n and h, which we suppress in the notation. For sake of brevity write

$$F(\mathbf{D}_{-\psi}(\mathbf{p}(x)/h+w\mathbf{e}_2-\mathbf{z}))=g_1(h\mathbf{z})\big\langle f(\mathbf{D}_{-\psi}(\mathbf{p}(x)/h+w\mathbf{e}_2-\mathbf{z}))g_2(h\mathbf{z}),g_3(h\mathbf{z})\big\rangle,$$

such that $Z_{n,G}(x,w,\psi)=\int_{\mathbb{R}^2}F(\mathbf{D}_{-\psi}(\mathbf{p}(x)/h+w\mathbf{e}_2-\mathbf{z}))\mathrm{d}W(\mathbf{z})$. Recall that F has compact support since $\mathrm{supp}(f)=[-1,1]$ and in addition F is Lipschitz continuous on its support with uniform Lipschitz constant $L_F=O(h^{r_2+r_3})$, compare to (A.4). Furthermore, the function

$$G(x,w,\psi):=\mathbf{D}_{-\psi}(\mathbf{p}(x)/h+w\mathbf{e}_2-\mathbf{z})$$

is Lipschitz continuous on Θ_n with Lipschitz constant $L_G=O(h^{-1})$, which is also uniform in $\mathbf{z}$. Hence, by Ito-Isometry and the Lipschitz continuity of F and G,

$$\begin{aligned}&\rho^2\big((x_1,w_1,\psi_1)^T,(x_2,w_2,\psi_2)^T\big)\\&=\frac{1}{h^{r_1-1}}\int_{A_1\cup A_2}|F(\mathbf{D}_{-\psi_1}(\mathbf{p}(x_1)/h+w_1\mathbf{e}_2-\mathbf{z}))-F(\mathbf{D}_{-\psi_2}(\mathbf{p}(x_2)/h+w_2\mathbf{e}_2-\mathbf{z}))|^2\,\mathrm{d}\mathbf{z}\\&\le\frac{1}{h^{r_1-1}}L_F^2L_G^2\,||(x_1,w_1,\psi_1)^T-(x_2,w_2,\psi_2)^T\,||^2\,\lambda_2(A_1\cup A_2),\end{aligned}$$

where A_i denotes the bounded set $\left\{\mathbf{z} \in \mathbb{R}^2 : \; F(\mathbf{D}_{-\psi_i}(\mathbf{p}(x_i)/h + w_i\mathbf{e}_2 - \mathbf{z})) \neq 0\right\}$ for $i = 1,2$ respectively. Furthermore, the semi-metric is bounded. Indeed, it easily follows since by choice $1 - r_1 + r_2 + r_3 = 0$ and because F is bounded that $\rho^2\big((x_1,w_1,\psi_1)^T,(x_2,w_2,\psi_2)^T\big) \le 2h^{1-r_1}\|F\|_\infty^2 \lambda_2(A_1 \cup A_2) \le 8C_F$, for some constant $C_F > 0$. Thus, for any $(x_1,w_1,\psi_1),(x_2,w_2,\psi_2) \in \Theta_n$,

$$\rho\big((x_1,w_1,\psi_1)^T,(x_2,w_2,\psi_2)^T\big) \le \sqrt{8}\,C_F \wedge Ch^{-1}\|(x_1,w_1,\psi_1)^T - (x_2,w_2,\psi_2)^T\|, \tag{A.7}$$

for some constant $C > 0$ uniform for x,w and ψ. With this, the diameter of Θ_n with respect to ρ is bounded by $diam_\rho(\Theta_n) \le \sqrt{8}\,C_F$ and in addition, the number of balls of radius $r > 0$ in the semi-metric ρ that cover Θ_n is not larger than $Cr^{-3}h^{-1}\lambda_3(\Theta_n)$. Moreover, $\sup_{(x,w,\psi)^T \in \Theta_n} \mathbb{E}|Z_{n,G}(x,w,\psi)|^2 \le C_0$ for some appropriate constant $C_0 > 0$, which can be chosen uniformly in x,w and ψ. Applying Proposition A.2.7 in van der Vaart and Wellner (1996) provides the first part of the lemma.

Second part: Order of the moment

For the second part use the second statement of Corollary 2.2.8 in van der Vaart and Wellner (1996) and Lemma A.5 together with a substitution to see that

$$\begin{aligned}\mathbb{E}\Big(\sup_{(x,w,\psi)\in\Theta_n} |Z_{n,G}(x,0,\psi(x))|\Big) &\le \mathbb{E}|Z_{n,G}(0,0,\psi(x))| + C_1\int_0^{\sqrt{8}C_F} \sqrt{\log\left(Cy^{-1}h^{-1}\right)}\,\mathrm{d}y \\ &\le C_2 + C_3\sqrt{\log\left(n\right)},\end{aligned}$$

where $C_i > 0$ are appropriate constants for $i = 1,2,3$, which can be chosen uniformly for x,w and ψ. Note that for the last line of the preceding display we used that $\sqrt{\log\left(n\right)} \cong \sqrt{\log\left(h^{-1}\right)}$, which is implied by the preconditions of this lemma.

Third part: Order of the increments

From (A.7) derive that

$$\rho\big((x_1,w_1,\psi_1)^T,(x_2,w_2,\psi_2)^T\big) \le Ch^{-1}\|(x_1,w_1,\psi_1)^T - (x_2,w_2,\psi_2)^T\|,$$

for some finite constant $C > 0$ uniform for x,w and ψ. The first statement of Corollary 2.2.8 in van der Vaart and Wellner (1996) and leads to

$$\begin{aligned}&\mathbb{E}\sup_{\theta_1,\theta_2\in\Theta_n\,:\|\theta_1-\theta_2\|\le\delta} |Z_{n,G}(\theta_1) - Z_{n,G}(\theta_2)| \\ &\le \mathbb{E}\sup_{\theta_1,\theta_2\in\Theta_n\,:\rho(\theta_1,\theta_2)\le Ch^{-1}\delta} |Z_{n,G}(\theta_1) - Z_{n,G}(\theta_2)| \\ &\le C_4\int_0^{Ch^{-1}\delta} \sqrt{\log\left(Cy^{-1}h^{-1}\right)}\,\mathrm{d}y\end{aligned}$$

where C_4 is some finite absolute constant uniform in x,w and ψ. Proceeding similarly as in the second part yields the assertion.

Fourth part: Order of the quantile

With the second part of this proof obtain $\mathbb{E}\Big(\sup_{(x,w,\psi)\in\Theta_n} |Z_{n,G}(x,w,\psi)|\Big) \le C\sqrt{\log(n)}$ for some constant $C > 0$. By Sudakov's inequality (see Proposition A.2.5 in van der Vaart and Wellner

(1996)),

$$c\sqrt{\log(n)} \leq \mathbb{E}\Big(\sup_{(x,w,\psi)\in\Theta_n} |Z_{n,G}(x,w,\psi)|\Big)$$

for some constant $c > 0$. This concludes the fourth statement of the lemma in view of Lemma A.6. □

A.2 Properties of sub-Gaussian processes

Following Viens and Vizcarra (2007) we call a centered random variable ξ *sub-Gaussian relative to the scale* M, if for all $u > 0$

$$P(|\xi| > u) \leq 2\exp\Big(-\frac{3u^2}{M^2}\Big). \tag{A.8}$$

Let $(\xi_i)_{i=1,\dots,n}$ be an i.i.d. sequence of random variables such that $\mathbb{E}(\xi_1) = 0$, $\mathbb{E}(\xi_1^2) = \sigma^2$ and ξ_1 is sub-Gaussian relative to the scale σ_g with $\sigma_g \geq \sigma > 0$. Define the process

$$Z_n(t;h) := (nh)^{-1/2}\sum_{i=1}^{n}\xi_i K(h^{-1}(x_i - t)), \quad t \in [0,1]$$

where $h < 1$ and $K : \mathbb{R} \to \mathbb{R}$ is a function with the following properties:

1. K is bounded and Lipschitz continuous with Lipschitz constant $C_K > 0$.
2. $\operatorname{supp}(K) = [-1,1]$.

Lemma A.7. *There exist constants $C_1, C_2, h_0 > 0$ depending only on K and σ_g, such that for any $\lambda > 0$ and $h \in (0,h_0)$ such that $\lambda > C_1\sqrt{-\log(h)}$, it holds that*

$$P\left(\sup_{t\in[0,1]} |Z_n(t;h)| \geq \lambda\right) \leq 2\exp\big(-C_2\lambda^2\big).$$

Lemma A.8. *There exist constants $C, h_0 > 0$ depending only on K and σ_g, such that if $h \in (0,h_0)$ then*

$$\mathbb{E}\sup_{t\in[0,1]} |Z_n(t;h)| \leq C\sqrt{\log(1/h)}.$$

To prove Lemma A.7 resp. Lemma A.8 we make use of Theorem 3.1 resp. Corollary 3.3 in Viens and Vizcarra (2007), of which we derive the requirements in the following. Define the semi-metric $\rho_{n,h} : [0,1]^2 \to \mathbb{R}_+$ by $\rho_{n,h}^2(s,t) = \mathbb{E}(Z_n(s;h) - Z_n(t;h))^2$. In the following we suppress the dependency of $\rho_{n,h}$ on n in the notation and just write ρ_h. This is due to our upper bound in (A.13) below. We write $N(\rho_h, T, \varepsilon)$ to denote the smallest number of ρ_h-balls of radius ε needed to cover $T \subset (0,1)$.

Lemma A.9. *(i) There exists a constant $c_g > 0$ depending only on K and σ_g such that for any $s,t \in [0,1]$ the random variable $c_g\big(Z_n(t;h) - Z_n(s;h)\big)$ is sub-Gaussian relative to the scale $\rho_h(s,t)$.*

(ii) $diam_{\rho_h}[0,1] \leq 2\sigma_g^2 ||K||_\infty^2$.

(iii) For any $T \subset [0,1]$, there exist some finite constants $C, h_0 > 0$ depending only on K and σ_g such that if $h \in (0,h_0)$

$$\int_0^\infty \sqrt{\log(N(\rho_h, T, x))}\,\mathrm{d}x \leq C\sqrt{\log(1/h)}.$$

Proof of Lemma A.7. First, note by the triangle inequality that for any $t_0 \in [0,1]$ holds

$$P\left(\sup_{t\in[0,1]} |Z_n(t;h)| \geq \lambda\right) \leq P\left(\sup_{t\in[0,1]} |Z_n(t;h) - Z_n(t_0;h)| \geq \lambda/2\right) + P\left(|Z_n(t_0;h)| \geq \lambda/2\right). \quad \text{(A.9)}$$

Since $Z_n(t_0;h)$ is the sum of independent sub-Gaussian random variables, we can apply the general Hoeffding inequality (see for instance Theorem 2.6.3 in Vershynin (2018)) to obtain

$$P(|Z_n(t_0;h)| \geq \lambda/2) \leq 2\exp(-\tilde{C}\lambda^2), \quad \text{(A.10)}$$

where $\tilde{C} > 0$ is a finite constant depending only on K and σ_g. Next, the process $Z = (c_g\, Z_n(t;h))_{t\in[0,1]}$ is separable and by Lemma A.9, (i), a sub-1th-Gaussian chaos field (see Definition 2.3 in Viens and Vizcarra (2007)) with respect to ρ. Without loss of generality let us assume that the constant c_g in Lemma A.9, (i), is one. Otherwise, we consider the random process $\tilde{Z}_n(t,h) = c_g^{-1} Z_n(t,h)$ and incorporate the constant c_g within the constants C_1 and C_2. Thus, by Theorem 3.1 in Viens and Vizcarra (2007) for any $\lambda > \tilde{C}_1 M$, where $\tilde{C}_1 > 0$ is a finite constant depending only on K and σ_g and $M = \int_0^\infty \sqrt{\log(N(\rho_h,T,x))}\,\mathrm{d}x$ it holds for a suitable finite constant $\tilde{C}_2 > 0$ depending only on K and σ_g that

$$P\left(\sup_{t\in[0,1]} |Z_n(t;h) - Z_n(t_0;h)| \geq \lambda/2\right) \leq 2\exp\left(-\tilde{C}_2\lambda^2\right), \quad \forall t_0 \in [0,1]. \quad \text{(A.11)}$$

In view, of Lemma A.9, (iii), $M \leq C\sqrt{\log(1/h)}$ if $h < h_0$, where $h_0, C > 0$ are finite constants depending only on K and σ_g. Hence, choose $C_1 := C\tilde{C}_1$ and $C_2 := \tilde{C} + \tilde{C}_2$ to conclude the proof, due to (A.9), (A.10) and (A.11). □

Proof of Lemma A.8. As in the proof of Lemma A.7 we can assume without loss of generality that the constant c_g in Lemma A.9, (i), is one. Using Corollary 3.4. in Viens and Vizcarra (2007) yields the assertion by using the bound on the covering entropy in 3. of Lemma A.9. □

Proof of Lemma A.9. Ad(i).

It holds that

$$\begin{aligned}\rho_h(s,t)^2 &= \mathbb{E}|Z_n(s;h) - Z_n(t;h)|^2 \\ &= \sigma^2 (nh)^{-1} \sum_{i=1}^n |K(h^{-1}(x_i - t)) - K(h^{-1}(x_i - s))|^2.\end{aligned} \quad \text{(A.12)}$$

Let $s,t \in [0,1]$ and $u > 0$. By means of the general Hoeffding inequality for sums of independent sub-Gaussian random variables (see for instance Theorem 2.6.3 in Vershynin (2018)) obtain

$$P(|Z_n(s;h) - Z_n(t;h)| > u) \leq 2\exp\left(-C\frac{u^2}{\rho_h(s,t)^2}\right),$$

where $C > 0$ is some finite constant depending only on K and σ_g. Therefore, choosing $c_g > 0$ appropriately and depending only on K and σ_g we can observe from the latter display that

$$P(c_g\,|Z_n(s;h) - Z_n(t;h)| > u) \leq 2\exp\left(-3\frac{u^2}{\rho_h(s,t)^2}\right).$$

This shows *(i)* in view of (A.8).

Ad(ii).

Now, the right-hand side of (A.12) can be bounded in two ways. On the one hand, let $A(n,s,t)$ denote the set of indices for which the latter sum is not zero. Due to the compact support of K and the design assumption we have that $|A(n,s,t)| \leq Cnh$, for some finite constant $C > 0$. Thus, with the Lipschitz continuity of K

$$\rho_h(s,t)^2 \leq h^{-2} C C_K^2 |t-s|^2 \sigma_g^2,$$

since $\sigma_g \geq \sigma$. On the other hand, since K is bounded and due to the cardinality of $A(n,s,t)$ one has that $\rho_h(s,t)^2 \leq 2\sigma_g^2 ||K||_\infty^2$. Hence,

$$\rho_h(s,t)^2 \leq 2\sigma_g^2 ||K||_\infty^2 \wedge h^{-2} C^2 C_K^2 |t-s|^2 \sigma_g^2, \quad \forall s,t \in [0,1]. \tag{A.13}$$

This yields $diam_{\rho_h}[0,1] \leq 2\sigma_g^2 ||K||_\infty^2$.

Ad (iii).

Since the latter display relates the ρ-distance of s and t to their absolute distance it follows that for any $\varepsilon \in (0, diam_{\rho_h}[0,1])$ and any $T \subset [0,1]$ one has $N(\rho_h, T, \varepsilon) \leq C_1 \lambda_1(T)(h\varepsilon)^{-1}$ for some appropriate constant $C_1 > 0$ depending only on K and σ_g. With this and if $h_0 > 0$ is chosen appropriately small depending on $diam_{\rho_h}[0,1]$ deduce for any $h \in (0, h_0)$

$$\begin{aligned}
\int_0^\infty \sqrt{\log(N(\rho_h,T,x))}\,\mathrm{d}x &= \int_0^{diam_{\rho_h}[0,1]} \sqrt{\log(N(\rho_h,T,x))}\,\mathrm{d}x \\
&\leq \sqrt{\log(1/h)}\, diam_{\rho_h}[0,1] + \int_0^{diam_{\rho_h}[0,1]} \sqrt{\log(C_1 \lambda_1(T) x^{-1})}\,\mathrm{d}x \\
&\leq C_2 \sqrt{\log(1/h)},
\end{aligned}$$

for some finite constant $C_2 > 0$ depending only on K and σ_g. In view of (ii), the choice of h_0 depends only on K and σ_g as well which concludes the proof. □

APPENDIX B

Asymptotics of components of the contrast function and their derivatives

B.1 Bivariate design

We consider in this section the same setting as in Section A.1. In addition let $g_4 : \mathbb{R}^2 \to \mathbb{R}^2$ resp. $\tilde{g}_4 : \mathbb{R}^2 \to \mathbb{R}^2$ be functions with analogous shapes as g_2, g_3 resp. $\tilde{g}_2$, $\tilde{g}_3$.

In order to control the smooth part of the regression function in model (2.1) we will often use the following lemma. The proof is given below.

Lemma B.1. *Let $r_1 = 2 + j(r_2 + r_3)$ and $j \in \{1,2\}$, then it holds that*

$$\begin{aligned} S_n(x;w,\psi) &:= \left(n^2 h^{r_1}\right)^{-1} \sum_{i_1,i_2=1}^{n} g_1(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}) \left\langle f\left(h^{-1}\mathbf{D}_{-\psi}\left(\mathbf{p}_{w,h}(x) - \mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}\right)\right) g_2(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}), g_3(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}) \right\rangle^j \\ &= \int_{[-1,1]^2} g_1(\mathbf{p}_{w,h}(x) - h\mathbf{D}_{\psi}\mathbf{z}) \langle f(\mathbf{z})\tilde{g}_2(\mathbf{z}), \tilde{g}_3(\mathbf{z})\rangle^j \, \mathrm{d}\mathbf{z} + O\left(\left(nh^{r_1-1-r_2-r_3}\right)^{-1}\right) \\ &= g_1(\mathbf{p}_{w,h}(x)) \int_{[-1,1]^2} \langle f(\mathbf{z})\tilde{g}_2(\mathbf{z}), \tilde{g}_3(\mathbf{z})\rangle^j \, \mathrm{d}\mathbf{z} + O(h) + O\left(\left(nh^{r_1-1-r_2-r_3}\right)^{-1}\right), \end{aligned}$$

uniformly for x, w and ψ.

More generally, let $r_1 = 2 + 2r_2 + r_3 + r_4$. Then,

$$\begin{aligned} \tilde{S}_n(x;w,\psi) &:= \left(n^2 h^{r_1}\right)^{-1} \sum_{i_1,i_2=1}^{n} g_1\left(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}\right) \left\langle f(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x) - \mathbf{x}_{\mathbf{i_1},\mathbf{i_2}})) g_2(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}), g_3\left(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}\right) \right\rangle \\ &\qquad\qquad \times \quad \left\langle f(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x) - \mathbf{x}_{\mathbf{i_1},\mathbf{i_2}})) g_2(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}), g_4\left(\mathbf{x}_{\mathbf{i_1},\mathbf{i_2}}\right) \right\rangle \\ &= g_1\left(\mathbf{p}_{w,h}(x)\right) \int_{[-1,1]^2} \langle f(\mathbf{z})\tilde{g}_2(\mathbf{z}), \tilde{g}_3(\mathbf{z})\rangle \langle f(\mathbf{z})\tilde{g}_2(\mathbf{z}), \tilde{g}_4(\mathbf{z})\rangle \, \mathrm{d}\mathbf{z} + O(h) \\ &\qquad + O\left(\left(nh^{r_1-1-r_2-r_3}\right)^{-1}\right) + O\left(\left(nh^{r_1-1-r_2-r_4}\right)^{-1}\right), \end{aligned}$$

uniformly for x, w and ψ.

To take care of the jump part of the image in model (2.1), i.e. $j_\tau(\mathbf{z}) = \tau(z_1) 1_{[0,\phi(z_1)]}(z_2)$ for $\mathbf{z} = (z_1, z_2)^T \in [-1,1]^2$, we need the following lemma. The proof is given below.

Lemma B.2. *Let $r_1 = 2 + r_2 + r_3$. Suppose that $\phi \in C^2[0,1]$ and $\tau \in C^2(\mathbb{R}_+)$, then it follows that*

$$\begin{aligned} J_n(x;w,\psi) &:= \left(n^2 h^{r_1}\right)^{-1} \sum_{i_1,i_2=1}^{n} j_\tau(\mathbf{x_{i_1,i_2}}) \left\langle f(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x) - \mathbf{x_{i_1,i_2}}))g_2(\mathbf{x_{i_1,i_2}}), g_3(\mathbf{x_{i_1,i_2}}) \right\rangle \\ &= \int_{h^{-1}\mathbf{D}_{-\psi}\left(\mathbf{p}(x) - [0,1]^2 \setminus epi(\phi)\right) + w(\sin(\psi),\cos(\psi))^T} \tau\left(x - (h\mathbf{D}_\psi \mathbf{z})_1\right) \langle f(\mathbf{z})\tilde{g}_2(\mathbf{z}), \tilde{g}_3(\mathbf{z}) \rangle \, \mathrm{d}\mathbf{z} \\ &\quad + O\left((nh)^{-1}\right) \\ &= \tau(x) \int_{\mathbb{H}(\psi(x)-\psi) + w(\sin(\psi),\cos(\psi))^T} \langle f(\mathbf{z})\tilde{g}_2(\mathbf{z}), \tilde{g}_3(\mathbf{z}) \rangle \, \mathrm{d}\mathbf{z} + O(h) + O\left((nh)^{-1}\right), \end{aligned}$$

uniformly for x, w and ψ, where $\mathbb{H}(\psi) = \mathbf{D}_\psi\left(\mathbb{R} \times [0,\infty)\right)$, as defined in (2.17).

For handling the error terms of the empirical contrast function and its derivatives, we formulate the following result.

Lemma B.3. *Let $r_1 = 2 + r_2 + r_3$ then under the assumptions of Lemma A.1, one obtains if n is large enough*

$$E_n(x;w,\psi) := \left(n^2 h^{r_1}\right)^{-1} \sum_{i_1,i_2=1}^{n} \boldsymbol{\varepsilon}_{i_1,i_2} \left\langle f(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x) - \mathbf{x_{i_1,i_2}}))g_2(\mathbf{x_{i_1,i_2}}), g_3(\mathbf{x_{i_1,i_2}}) \right\rangle = o_P(h),$$

uniformly for x, w and ψ.

Proof. Note that ${}^{nh}/_{\sigma} E_n(x;w,\psi) = Z_{n,0}(x,w,\psi)$, where $Z_{n,0}$ is given in (A.2). Moreover, note that the exponent of the bandwidth in $Z_{n,0}$ is now $r_1 - 1$, which satisfies the assumption on the exponent in (A.2), due to the precondition of this lemma. Moreover, ${}^{nh^2}/_{2\sigma} > C_f$ for sufficiently large n. Hence, Lemma A.1 and part one of Lemma A.3 imply

$$\begin{aligned} P\Big(\sup_{(x,w,\psi)^T \in \Theta_n} |E_n(x,w,\psi)| > h\Big) &\le P\Big(\sup_{(x,w,\psi)^T \in \Theta_n} |Z_{n,G}(x,w,\psi)| > {}^{nh^2}/_{2\sigma}\Big) + P\Big(R_n > {}^{nh^2}/_{2\sigma}\Big) \\ &\le C_3 \frac{n}{h^3} \exp\left(-C_4 (nh^2)^2\right) + o(1) = o(1), \end{aligned}$$

where $C_3, C_4 > 0$ are appropriate and finite constants and $R_n = O_P\Big(\sqrt{\log(n)}/_{n^{1/2}h}\Big)$, which is the approximation term in Lemma A.1. □

Proofs of Lemmas B.1 and B.2

We use $C > 0$ as a generic constant which can vary at every appearance.

Proof of Lemma B.1. We only prove the first representation, as the second can be derived analogously. For sake of brevity let us write S_n for $S_n(x;w,\psi)$ and

$$\check{F}(\mathbf{z};x,w,\psi,h) = \left\langle f\left(h^{-1}\mathbf{D}_{-\psi}\left(\mathbf{p}_{w,h}(x) - \mathbf{z}\right)\right) g_2(\mathbf{z}), g_3(\mathbf{z}) \right\rangle.$$

By Riemann-sum approximation

$$S_n = h^{-r_1} \int g_1(\mathbf{z}) \check{F}(\mathbf{z};x,w,\psi,h)^j \, \mathrm{d}\mathbf{z} + O\left(\left(nh^{r_1 - 1 - r_2 - r_3}\right)^{-1}\right),$$

uniformly for x,w and ψ. To see this, write $B(n,h)$ for the index set for which the sum in S_n is not zero and notice that $|B(n,h)| \leq 4n^2h^2$, due to the equidistant design and due to $\operatorname{supp}(f) = [-1,1]$. Let

$$A_{i_1,i_2} = [x_{i_1}, x_{i_1+1}) \times [x_{i_2}, x_{i_2+1}), \quad i_1, i_2 \in \{1, \ldots, n-1\},$$

where the right boundary is included if $i_1 = n-1$ or $i_2 = n-1$. Notice that $|x^2 - y^2| \leq 2C|x-y|$ for $x, y \in A$, where A is a compact subset in $\mathbb{R}$ and $C = \sup A$. Thus, for any $\mathbf{y}_1, \mathbf{y}_2 \in [0,1]^2$,

$$|g_1(\mathbf{y}_1)\check{F}(\mathbf{y}_1; x, w, \psi, h)^2 - g_1(\mathbf{y}_2)\check{F}(\mathbf{y}_2; x, w, \psi, h)^2| \leq C|\sqrt{g_1(\mathbf{y}_1)}\check{F}(\mathbf{y}_1; x, w, \psi, h) - \sqrt{g_1(\mathbf{y}_2)}\check{F}(\mathbf{y}_2; x, w, \psi, h)|,$$

for some suitable constant $C > 0$. So it suffices to consider the case $j = 1$ by controlling the error term, say Err, in the Riemann-sum approximation. As the product of g_1 and $\check{F}$ is a C^1-function, the mean value theorem leads to

$$\begin{aligned} Err &\leq \left(n^2h^{r_1}\right)^{-1} \sum_{i_1,i_2 \in B(n,h)} |\sup_{\mathbf{y}_1 \in A_{i_1,i_2}} g_1(\mathbf{y}_1)\check{F}(\mathbf{y}_1; x, w, \psi, h) - \inf_{\mathbf{y}_2 \in A_{i_1,i_2}} g_1(\mathbf{y}_2)\check{F}(\mathbf{y}_2; x, w, \psi, h)| \\ &\leq \frac{n^2h^2}{n^3h^{r_1}} \sup_{i_1,i_2 \in B(n,h)} \sup_{\mathbf{y} \in A_{i_1,i_2}} ||\nabla g_1(\mathbf{y})\check{F}(\mathbf{y}; x, w, \psi, h)||, \end{aligned}$$

since by the equidistant design $\sup_{i_1,i_2 \in B(n,h)} \sup_{\mathbf{y}_1,\mathbf{y}_2 \in A_{i_1,i_2}} ||\mathbf{y}_1 - \mathbf{y}_2|| \leq n^{-1}$. Considering the special representations of g_2 and g_3 we obtain by the chain rule

$$\sup_{i_1,i_2 \in B(n,h)} \sup_{\mathbf{y} \in A_{i_1,i_2}} ||\nabla g_1(\mathbf{y})\check{F}(\mathbf{y}; x, w, \psi, h)|| = O(h^{r_2+r_3-1}),$$

uniformly for x,w and ψ. This uniformity comes from the fact that x and ψ are elements of compact sets and w is such that $\mathbf{p}_{w,h}(x)$ is element of a compact subset. Therefore, $Err = O\big((nh^{r_1-1-r_2-r_3})^{-1}\big)$ uniformly over x,w and ψ as claimed. With the substitution

$$\mathbf{z} \mapsto h^{-1}\mathbf{D}_{-\psi}\left(\mathbf{p}_{w,h}(x) - \mathbf{z}\right) \tag{B.1}$$

one has

$$\begin{aligned} S_n &= h^{2-r_1} \int_{[-1,1]^2} g_1\left(\mathbf{p}_{w,h}(x) - h\mathbf{D}_\psi \mathbf{z}\right) \check{F}(\mathbf{p}_{w,h}(x) - h\mathbf{D}_\psi \mathbf{z}; x, w, \psi, h)^j \,\mathrm{d}\mathbf{z} + O\big((nh^{r_1-1-r_2-r_3})^{-1}\big) \\ &= \int_{[-1,1]^2} g_1(\mathbf{p}_{w,h}(x) - h\mathbf{D}_\psi \mathbf{z}) \langle f(\mathbf{z})\tilde{g}_2(\mathbf{z}), \tilde{g}_3(\mathbf{z}) \rangle^j \,\mathrm{d}\mathbf{z} + O\big((nh^{r_1-1-r_2-r_3})^{-1}\big), \end{aligned}$$

uniformly for x,w and ψ, where we used the choice of r_1 as well as (A.1) in the last line. Additionally, using the smoothness of g_1 obtain for $\mathbf{z} \in [-1,1]^2$ that

$$|g_1(\mathbf{p}_{w,h}(x) - h\mathbf{D}_\psi \mathbf{z}) - g_1(\mathbf{p}_{w,h}(x))| \leq Ch,$$

where $C > 0$ is a constant, which can be chosen uniformly for x, w and ψ as well. Hence, from the previous two displays one derives

$$S_n = g_1\left(\mathbf{p}_{w,h}(x)\right) \int_{[-1,1]^2} \langle f(\mathbf{z})\tilde{g}_2(\mathbf{z}), \tilde{g}_3(\mathbf{z}) \rangle^j \,\mathrm{d}\mathbf{z} + O(h) + O\big((nh^{r_1-1-r_2-r_3})^{-1}\big),$$

uniformly for x,w and ψ. □

Proof of Lemma B.2. For sake of brevity write J_n for $J_n(x; w, \psi)$ and

$$G(\mathbf{z}; x, w, \psi, h) := j_\tau(\mathbf{z}) \left\langle f(h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x) - \mathbf{z}))g_2(\mathbf{z}), g_3(\mathbf{z}) \right\rangle.$$

Let A_{i_1,i_2} and $B(n,h)$ be as in the proof of Lemma B.1. Additionally, let $E(n,h)$ denote the set of indices in the sum of J_n, for which the design points intersect with the curve $y=\phi(x)$ and $B^*(n,h)=B(n,h)\setminus E(n,h)$ be the set of the remaining indices for which the sum is not zero. Notice that $|E(n,h)|=O(nh)$, due to smoothness of ϕ and $|B^*(n,h)|\leq|B(n,h)|\leq 4n^2h^2$. By Riemann-sum approximation

$$J_n=h^{-r_1}\int G(\mathbf{z};x,w,\psi,h)\,\mathrm{d}\mathbf{z}+O\big((nh)^{-1}\big),$$

uniformly for x,w and ψ. Indeed, the error term Err in the Riemann-sum approximation can be bounded as follows

$$\begin{aligned}
Err &\leq \left(n^2h^{r_1}\right)^{-1}\sum_{i_1,i_2\in B^*(n,h)}\Big|\sup_{\mathbf{y}_1\in A_{i_1,i_2}}G(\mathbf{y}_1;x,w,\psi,h)-\inf_{\mathbf{y}_2\in A_{i_1,i_2}}G(\mathbf{y}_2;x,w,\psi,h)\Big|\\
&\quad+\left(n^2h^{r_1}\right)^{-1}\sum_{i_1,i_2\in E(n,h)}\Big|\sup_{\mathbf{y}_1\in A_{i_1,i_2}}G(\mathbf{y}_1;x,w,\psi,h)-\inf_{\mathbf{y}_2\in A_{i_1,i_2}}G(\mathbf{y}_2;x,w,\psi,h)\Big|\\
&=: Err_{(1)}+Err_{(2)}.
\end{aligned}$$

Note that $G(\cdot;x,w,\psi,h)$ is a C^1 function on the design squares A_{i_1,i_2} for $i_1,i_2\in B^*(n,h)$. Hence, we can proceed for $Err_{(1)}$ as in the proof of Lemma B.1 to derive that $Err_{(1)}\leq C(nh)^{-1}$, due to choice of r_1, where the constant $C>0$ can be chosen uniformly for x, w and ψ. Additionally, by the special form of g_2 and g_3

$$\sup_{i_1,i_2\in E(n,h)}\sup_{\mathbf{y}\in A_{i_1,i_2}}||g_i(\mathbf{y})||\leq Ch^{r_i},\quad i=2,3,$$

where the constant $C>0$ can be chosen uniformly for x, w and ψ. With this it easily follows, $Err_{(2)}\leq C(nh)^{-1}$ such that $Err\leq C(nh)^{-1}$ for some constant $C>0$ which is uniform in x, w and ψ. Let

$$\mathbb{H}_{x,\psi,h}=h^{-1}\mathbf{D}_{-\psi}\left(\mathbf{p}(x)-[0,1]^2\backslash\mathrm{epi}(\phi)\right).$$

With the substitution (B.1) and (A.1) obtain

$$\begin{aligned}
J_n&=h^{2-r_1}\int_{h^{-1}\mathbf{D}_{-\psi}(\mathbf{p}_{w,h}(x)-[0,1]^2\backslash\mathrm{epi}(\phi))}\tau\big(x-(h\mathbf{D}_\psi\mathbf{z})_1\big)G(\mathbf{p}_{w,h}(x)-h\mathbf{D}_\psi\mathbf{z};x,w,\psi,h)\,\mathrm{d}\mathbf{z}+O\big((nh^{r_1-1-r_2-r_3})^{-1}\big)\\
&=\int_{\mathbb{H}_{x,\psi,h}+w(\sin(\psi),\cos(\psi))^T}\tau\big(x-(h\mathbf{D}_\psi\mathbf{z})_1\big)\,\langle f(\mathbf{z})\tilde g_2(\mathbf{z}),\tilde g_3(\mathbf{z})\rangle\,\mathrm{d}\mathbf{z}+O\big((nh^{r_1-1-r_2-r_3})^{-1}\big)\\
&=\tau(x)\int_{\mathbb{H}(\psi(x)-\psi)+w(\sin(\psi),\cos(\psi))^T}\langle f(\mathbf{z})\tilde g_2(\mathbf{z}),\tilde g_3(\mathbf{z})\rangle\,\mathrm{d}\mathbf{z}+R_n+O\big((nh^{r_1-1-r_2-r_3})^{-1}\big),
\end{aligned}$$

uniformly for x,w and ψ, where

$$\begin{aligned}
R_n&=\int_{\mathbb{H}_{x,\psi,h}+w(\sin(\psi),\cos(\psi))^T}\left[\tau\big(x-(h\mathbf{D}_\psi\mathbf{z})_1\big)-\tau(x)\right]\langle f(\mathbf{z})\tilde g_2(\mathbf{z}),\tilde g_3(\mathbf{z})\rangle\,\mathrm{d}\mathbf{z}\\
&\quad+\int_{\mathbb{H}(\psi(x)-\psi)+w(\sin(\psi),\cos(\psi))^T\,\triangle\,\mathbb{H}_{x,\psi,h}+w(\sin(\psi),\cos(\psi))^T}\tau(x)\langle f(\mathbf{z})\tilde g_2(\mathbf{z}),\tilde g_3(\mathbf{z})\rangle\,\mathrm{d}\mathbf{z}\\
&=: R_{n,1}+R_{n,2}.
\end{aligned}$$

By smoothness assumptions on τ and the compact support of f it holds $|R_{n,1}|\leq Ch$, where $C>0$ can be chosen uniformly for x, w and ψ. Moreover,

$$\lambda_2(\mathbb{H}(\psi(x)-\psi)\,\triangle\,\mathbb{H}_{x,\psi,h})\leq Ch^2$$

and $C>0$ is independent of w and can be chosen uniformly as x and ψ take values in a compact subset of $\mathbb{R}^2$. Taking the compact support of f into account, leads to

$$|R_{n,2}| \leq \sup_{\mathbf{z}\in[-1,1]^2} ||f(\mathbf{z})\cdot\tilde{g}_2(\mathbf{z})||\cdot||\tilde{g}_3(\mathbf{z})||\cdot||\tau(\mathbf{z})||\,\lambda_2(\mathbb{H}(\psi(x)-\psi)\,\triangle\,\mathbb{H}_{x,\psi,h}) = O(h^2),$$

uniformly for x, w and ψ. Therewith, $R_n = O(h)$ uniformly in x, w and ψ which completes the proof.

□

B.2 Univariate design

We consider in this section the model in Example 1 in Section 1.4. Furthermore, let $K : \mathbb{R} \to \mathbb{R}$ be two times continuous differentiable with $\operatorname{supp}(K) = [-1,1]$.

A more simpler version of Lemma B.1 is sufficient to take care of the discretization error-terms for the one-dimensional design case.

Lemma B.4. *Let $r \in \mathbb{N}$, and $g : [0,1] \to \mathbb{R}$ is Lipschitz-continuous. Then, for $h < 1/2$ small enough it holds for any $n \in \mathbb{N}$ and any $x \in I$ that*

$$\frac{1}{nh^r}\sum_{i=1}^{n} g(x_i)K(h^{-1}(x_i-x)) = h^{-r}\int g(y)K(h^{-1}(y-x))\,\mathrm{d}y + O_{x\in[0,1]}((nh^r)^{-1}).$$

The proof is analogous to Lemma B.1.

Similarly, a simpler version of Lemma B.3 is appropriate to cope with the stochastic terms in the univariate model.

Lemma B.5. *Let $(\varepsilon_i)_{i=1,\ldots,n}$ be such that $\mathbb{E}|\varepsilon_1|^5 < \infty$ and $h = n^{-\eta}$ with $\eta \in (0,1/2)$. Then setting*

$$E_n(x) = (nh^r)\sum_{i=1}^{n}\varepsilon_i\,K(h^{-1}(x_i-x)), \quad x \in [0,1], r \in \mathbb{N}$$

one has for n sufficiently large that $\sup_{x\in[h,1-h]}|E_n(x)| = O_P\left(\sqrt{\log(n)/nh^{2r-1}}\right)$.

Proof. Define for $x \in [0,1]$,

$$Z_n(x) := (nh)^{-1/2}\sum_{i=1}^{n}\varepsilon_i K(h^{-1}(x_i-x)), \quad \text{and} \quad Z_G(x) := h^{-1/2}\int_{\mathbb{R}} K(h^{-1}(x-t))\,dW(t)$$

and W is a Brownian motion on $\mathbb{R}$. Then by Lemma A.2 on a possibly enriched probability space for sufficiently large n it follows that

$$\sup_{x\in[h,1-h]}|Z_n(x)-Z_G(x)| = O_P\left(\sqrt{\frac{\log(n)}{nh}}\right).$$

Moreover, it is straightforward to show for the univariate process Z_G a similar result as part two of A.3, i.e. there exists a finite constant $C > 0$, which is uniform in $[0,1]$ such that

$$\mathbb{E}\sup_{x\in[h,1-h]}|Z_G(x)| \leq C\sqrt{\log(n)}.$$

Apparently $\sqrt{nh^{2r-1}}E_n = Z_n$ such that all things considered, for some finite constant $c_0 > 0$ obtain by Markov's inequality

$$\begin{aligned}
&P(\sup_{x\in[h,1-h]} E_n(x) > c_0\sqrt{\log(n)/nh^{2r-1}}) \\
&\qquad \le P(\sup_{x\in[h,1-h]} |Z_n(x) - Z_G(x)| > c_0\sqrt{\log(n)}/2) + P(\sup_{x\in[h,1-h]} |Z_G(x)| > c_0\sqrt{\log(n)}/2) \\
&\qquad \le o(1) + \frac{2\,C}{c_0},
\end{aligned}$$

which completes the proof by a suitable choice of c_0. □

APPENDIX C

Extended probability theory

We use the following notation in this section. By F_X we denote the cumulative distribution function of a random vector X and by P_X its law. Let ϕ_X be the characteristic function of a random vector X.

Moreover, we assume for this section that Θ is some arbitrary set and for any $\vartheta \in \Theta$, $(X_n^\vartheta)_{n\in\mathbb{N}}$ is a sequence of real-valued random vectors in $\mathbb{R}^d$. Likewise, for any $\vartheta \in \Theta$, let X^ϑ be random vectors in $\mathbb{R}^d$ with continuous distribution. We introduce some definitions for the remainder of this section.

Uniform convergence in distribution

We write $X_n^\vartheta \overset{D,\Theta}{\Longrightarrow} X^\vartheta$ if

$$\sup_{\vartheta\in\Theta} |F_{X_n^\vartheta}(\mathbf{x}) - F_{X^\vartheta}(\mathbf{x})| = o(1), \quad \forall \mathbf{x} \in \mathbb{R}^d.$$

In this case we say that X_n^ϑ converges uniformly over Θ in distribution to X^ϑ. We say that $(P_{X^\vartheta})_{\vartheta\in\Theta}$ is *uniformly absolutely continuous over* Θ *with respect to some continuous probability measure* Q, if for any $\varepsilon > 0$ there exists a $\delta > 0$ such that for any measurable $A \subset \mathbb{R}^d$ with $Q(A) < \delta$ one has that $\sup_{\vartheta\in\Theta} P_{X^\vartheta}(A) < \varepsilon$. Note that by continuous probability measure we mean that the measure of singletons is zero, i.e. $Q(\{x\}) = 0$ for any $x \in \mathbb{R}^d$.

Uniform weak convergence

Likewise, we define uniform weak convergence for probability measures. Let $(\mathcal{X}, d)$ be some metric space and let $\mathcal{A}$ be its Borel σ-algebra. Let for any $\vartheta \in \Theta$, $(\mu_n^\vartheta)_{n\in\mathbb{N}}$ be a sequence of probability measures on $(\mathcal{X},\mathcal{A})$. Similarly, for any $\vartheta \in \Theta$, let μ^ϑ be a probability measure on $(\mathcal{X},\mathcal{A})$. In the same manner as for the law of random vectors we define uniform absolute continuity over Θ for $(\mu^\vartheta)_{\vartheta\in\Theta}$, that is $(\mu^\vartheta)_{\vartheta\in\Theta}$ is *uniformly absolutely continuous over* Θ *with respect to some continuous probability measure* μ if for any $\varepsilon > 0$ there exists a $\delta > 0$ such that for any $A \in \mathcal{A}$ with $\mu(A) < \delta$ it follows that $\sup_{\vartheta\in\Theta} \mu^\vartheta(A) < \varepsilon$. Eventually, we say that μ_n^ϑ converges uniformly weakly over Θ to μ^ϑ and write $\mu_n^\vartheta \overset{w,\Theta}{\Longrightarrow} \mu^\vartheta$ if and only if

$$\sup_{\vartheta\in\Theta} \left| \int g \, \mathrm{d}\mu_n^\vartheta - \int g \, \mathrm{d}\mu^\vartheta \right| = o(1)$$

for any real-valued, bounded and continuous function $g : \mathcal{X} \to \mathbb{R}$.

Uniform Portmanteau lemma

The following uniform version of Portmanteau's Lemma will be of great importance for the proofs of the following results.

Lemma C.1 (Uniform Portmanteau). *Let $(\mu^\vartheta)_{\vartheta\in\Theta}$ be uniformly absolutely continuous over Θ with respect to some continuous probability measure μ. The following conditions are equivalent:*

1. $\mu_n^\vartheta \overset{w,\Theta}{\Longrightarrow} \mu^\vartheta$;
2. $\sup_{\vartheta\in\Theta} |\int g \,\mathrm{d}\mu_n^\vartheta - \int g\,\mathrm{d}\mu^\vartheta| = o(1)$ *for any real-valued, bounded and Lipschitz continuous function* $g : \mathcal{X} \to \mathbb{R}$;
3. $\limsup_n \sup_{\vartheta\in\Theta}(\mu_n^\vartheta(F) - \mu^\vartheta(F)) \le 0$ *for any closed set* $F \subset \mathcal{X}$;
4. $\liminf_n \inf_{\vartheta\in\Theta}(\mu_n^\vartheta(G) - \mu^\vartheta(G)) \ge 0$ *for any open set* $G \subset \mathcal{X}$;
5. $\limsup_n \sup_{\vartheta\in\Theta} \left(g\,\mathrm{d}\mu_n^\vartheta - \int g\,\mathrm{d}\mu^\vartheta\right) \le 0$ *for any real-valued, bounded and upper-semi-continuous function* $g : \mathcal{X} \to \mathbb{R}$;
6. $\liminf_n \inf_{\vartheta\in\Theta} \left(g\,\mathrm{d}\mu_n^\vartheta - \int g\,\mathrm{d}\mu^\vartheta\right) \ge 0$ *for any real-valued, bounded and lower-semi-continuous function* $g : \mathcal{X} \to \mathbb{R}$;
7. $\sup_{\vartheta\in\Theta} |\mu_n^\vartheta(A) - \mu^\vartheta(A)| = o(1)$ *for any set* $A \in \mathcal{A}$ *such that* $\sup_{\vartheta\in\Theta} \mu^\vartheta(\partial A) = 0$.

*Proof of Lemma C.*1. It is clear, that 1. implies 2. Now, we show that 2. implies 3. Let $F \subset \mathcal{X}$ be a closed set and for $k \in \mathbb{N}$ let

$$F_k = \{x \in \mathcal{X} \mid d(x,F) \le 1/k\},$$

where $d(x,F) = \inf\{d(x,y) \mid y \in F\}$. Additionally, for any $x \in \mathcal{X}$ let $g_{k,F}(x) = \max\{1 - kd(x,F), 0\}$. Then $g_{k,F}$ is Lipschitz continuous with $1_F \le g_{k,F} \le 1_{F_k}$. Thus,

$$\begin{aligned}
\limsup_n \sup_{\vartheta\in\Theta}(\mu_n^\vartheta(F) - \mu^\vartheta(F)) &= \limsup_n \sup_{\vartheta\in\Theta}\Big(\int 1_F \mathrm{d}\mu_n^\vartheta - \int 1_F \mathrm{d}\mu^\vartheta\Big) \\
&\le \limsup_n \sup_{\vartheta\in\Theta}\Big(\int g_{k,F} \mathrm{d}\mu_n^\vartheta - \int 1_F \mathrm{d}\mu^\vartheta\Big) \\
&\le \limsup_n \sup_{\vartheta\in\Theta}\Big(\int g_{k,F} \mathrm{d}\mu_n^\vartheta - \int g_{k,F} \mathrm{d}\mu^\vartheta\Big) \\
&\quad + \limsup_n \sup_{\vartheta\in\Theta}\Big(\int g_{k,F} \mathrm{d}\mu^\vartheta - \int 1_{F_k} \mathrm{d}\mu^\vartheta\Big) \\
&\quad + \sup_{\vartheta\in\Theta}\Big(\int 1_{F_k} \mathrm{d}\mu^\vartheta - \int 1_F \mathrm{d}\mu^\vartheta\Big).
\end{aligned}$$

Now, the first term on the right-hand side of the latter inequality is zero by assumption 2. The second is smaller than zero by construction of $g_{k,F}$. Concerning the third term, note that by continuity of measure $\mu(F_k \backslash F) \to 0$ for $k \to \infty$. Consequently, considering the limit $k \to \infty$ it follows by the uniform absolute continuity of $(\mu^\vartheta)_{\vartheta\in\Theta}$ with respect to μ and reasons of monotonicity that the third term tends to zero.

3. is equivalent to 4. by taking complements. Similarly, 5. and 6. are apparently equivalent. Moreover, 5. and 6. together imply 1. We prove that 3. implies 5. Let $g : \mathcal{X} \to \mathbb{R}$ be a bounded, upper-semi-continuous function with $a < g < b$ for some $a, b \in \mathbb{R}$. Then,

$$\sup_{\vartheta\in\Theta}\Big(\int g\,\mathrm{d}\mu_n^\vartheta - \int g\,\mathrm{d}\mu^\vartheta\Big) = \sup_{\vartheta\in\Theta} \int_a^b \big(\mu_n^\vartheta(g \ge x) - \mu^\vartheta(g \ge x)\big)\,\mathrm{d}x$$

$$\leq \int_a^b \sup_{\vartheta\in\Theta} \big(\mu_n^\vartheta(g \geq x) - \mu^\vartheta(g \geq x)\big)\,\mathrm{d}x.$$

Thus, by Fatou's lemma and by 3.

$$\limsup_n \sup_{\vartheta\in\Theta} \Big(\int g\,\mathrm{d}\mu_n^\vartheta - \int g\,\mathrm{d}\mu^\vartheta \Big) \leq \int_a^b \limsup_n \sup_{\vartheta\in\Theta} \big(\mu_n^\vartheta(g \geq x) - \mu^\vartheta(g \geq x)\big)\,\mathrm{d}x \leq 0,$$

which implies 5. It remains to incorporate 7. in the implication flow. Next, we show that 3. and 4. imply 7. and vice versa. Starting with 3. and 4. implies 7., we consider some $A \in \mathcal{A}$ such that $\sup_{\vartheta\in\Theta} \mu^\vartheta(\partial A) = 0$. Let A° denote its interior and $\bar{A}$ its closure. Clearly, $A^\circ \subset A \subset \bar{A}$ and A° is open and $\bar{A}$ is closed. Furthermore, $\mu^\vartheta(A^\circ) = \mu^\vartheta(\bar{A}) = \mu^\vartheta(A)$ for any $\vartheta \in \Theta$. Thus, by 3.

$$\limsup_n \sup_{\vartheta\in\Theta} (\mu_n^\vartheta(A) - \mu^\vartheta(A)) \leq \limsup_n \sup_{\vartheta\in\Theta} (\mu_n^\vartheta(\bar{A}) - \mu^\vartheta(\bar{A})) \leq 0$$

and by 4.

$$\liminf_n \inf_{\vartheta\in\Theta} (\mu_n^\vartheta(A) - \mu^\vartheta(A)) \geq \liminf_n \inf_{\vartheta\in\Theta} (\mu_n^\vartheta(A^\circ) - \mu^\vartheta(A^\circ)) \geq 0.$$

Both latter inequalities together imply 7. Now suppose 7. holds. Let $x \in \mathcal{X}$ and $F \subset \mathcal{X}$ be closed. Define for $r \geq 0$

$$B_F(r) = \{x \in \mathcal{X} \mid d(x,F) \leq r\}, \quad \text{and} \quad C_F(r) = \{x \in \mathcal{X} \mid d(x,F) = r\}.$$

Then, $(C_F(r))_{r\geq 0}$ is a partition of $\mathcal{X}$. Note that there exists a countable set $R \subset [0,\infty)$ such that $\mu(C_F(r)) = 0$ for $r \in [0,\infty)\backslash R$, otherwise we could contradict the measure properties of μ. It holds that $\partial B_F(r) \subset C_F(r)$ and thus $\mu(\partial B_F(r)) = 0$ for $r \in [0,\infty)\backslash R$. Due to uniform absolute continuity of $(\mu^\vartheta)_{\vartheta\in\Theta}$ with respect to μ one has that $\sup_{\vartheta\in\Theta} \mu^\vartheta(\partial B_F(r)) = 0$ for $r \in [0,\infty)\backslash R$. Hence, there exists a sequence $r_k \searrow 0$ such that

$$\sup_{\vartheta\in\Theta} \mu^\vartheta(\partial B_F(r_k)) = 0, \quad \forall k \in \mathbb{N}.$$

Since $F \subset B_F(r_k)$ and by 7.

$$\limsup_n \sup_{\vartheta\in\Theta} (\mu_n^\vartheta(F) - \mu^\vartheta(B_F(r_k))) \leq \limsup_n \sup_{\vartheta\in\Theta} (\mu_n^\vartheta(B_F(r_k) - \mu^\vartheta(B_F(r_k))) = 0, \quad \forall k \in \mathbb{N}.$$

Considering the limit $k \to \infty$ and noticing that by reasons of monotonicity the left-hand side of the latter display tends to $\limsup_n \sup_{\vartheta\in\Theta}(\mu_n^\vartheta(F) - \mu^\vartheta(F))$ concludes the lemma. □

Uniform weak convergence on generators

The following lemma shows that it suffices to show uniform weak convergence on some generator which is closed under finite intersections.

Lemma C.2. *Let $(\mu^\vartheta)_{\vartheta\in\Theta}$ be uniformly absolutely continuous over Θ with respect to some continuous probability measure μ and let $\tilde{\mathcal{B}}$ be a collection of open subsets which is closed under finite intersections. Furthermore, each open set of $\mathcal{X}$ can be represented as a countable union of elements in $\tilde{\mathcal{B}}$. Then,*

$$\sup_{\vartheta\in\Theta} |\mu_n^\vartheta(A) - \mu^\vartheta(A)| = o(1) \tag{C.1}$$

for all $A \in \tilde{\mathcal{B}}$ implies that $\mu_n^\vartheta \overset{w,\Theta}{\Longrightarrow} \mu^\vartheta$.

Proof of Lemma C.2. Let $G_1, G_2 \in \tilde{\mathcal{B}}$. By (C.1) and since $G_1 \cap G_2 \in \tilde{\mathcal{B}}$, we have

$$\begin{aligned}
&\sup_{\vartheta\in\Theta} |\mu_n^\vartheta(I_1 \cup I_2) - \mu^\vartheta(I_1 \cup I_2)| \\
&\quad \le \sup_{\vartheta\in\Theta} \Big[\big|\mu_n^\vartheta(I_1) - \mu^\vartheta(I_1)\big| + \big|\mu_n^\vartheta(I_2) - \mu^\vartheta(I_2)\big| + \big|\mu_n^\vartheta(I_1 \cap I_2) - \mu^\vartheta(I_1 \cap I_2)\big| \Big] = o(1).
\end{aligned}$$

Therefore, we can assume without loss of generality that $\tilde{\mathcal{B}}$ is closed under finite unions. Let $G \subset \mathcal{X}$ be open and let $A_i \in \tilde{\mathcal{B}}$ be such that $G = \cup_{i=1}^\infty A_i$. Due to continuity of measure, for each $\delta > 0$ there exists an integer $\tilde{N}$ such that $\mu(G) \le \mu(\cup_{i=1}^{\tilde{N}} A_i) + \delta$. By the uniform absolute continuity of $(\mu^\vartheta)_{\vartheta\in\Theta}$ over Θ with respect to μ there exists an integer N such that for any $\varepsilon > 0$

$$\mu^\vartheta(G) \le \mu^\vartheta(\cup_{i=1}^N A_i) + \varepsilon, \quad \forall \vartheta \in \Theta.$$

Therefore,

$$\begin{aligned}
\liminf_n \inf_{\vartheta\in\Theta} (\mu_n^\vartheta(G) - \mu^\vartheta(G) + \varepsilon) &\ge \liminf_n \inf_{\vartheta\in\Theta} \big(\mu_n^\vartheta(G) - \mu^\vartheta(\cup_{i=1}^N A_i)\big) \\
&\ge \liminf_n \inf_{\vartheta\in\Theta} \big(\mu_n^\vartheta(\cup_{i=1}^N A_i) - \mu^\vartheta(\cup_{i=1}^N A_i)\big) = 0,
\end{aligned}$$

where the last equation is due to (C.1), since $\cup_{i=1}^N A_i \in \tilde{\mathcal{B}}$. Letting $\varepsilon \to 0$ completes the proof by means of 3. of Lemma C.1. □

Relation between uniform weak convergence and uniform convergence in distribution

The following theorem relates the uniform weak convergence with the uniform convergence in distribution.

Theorem C.3. *Let $(\mu^\vartheta)_{\vartheta\in\Theta}$ be uniformly absolutely continuous over Θ with respect to some continuous probability measure μ. Let $\mathcal{X} = \mathbb{R}^d$ and thus $\mathcal{A}$ be the Borel-σ-algebra on $\mathbb{R}^d$. Setting*

$$F_{\vartheta,n}(\mathbf{x}) = \mu_n^\vartheta\big((-\infty, \mathbf{x}]\big), \quad \text{and} \quad F_\vartheta(\mathbf{x}) = \mu^\vartheta\big((-\infty, \mathbf{x}]\big), \quad \mathbf{x} \in \mathbb{R}^d.$$

The following two statements are equivalent:

1. $\mu_n^\vartheta \overset{w,\Theta}{\Longrightarrow} \mu^\vartheta$;
2. $\sup_{\vartheta\in\Theta} |F_{\vartheta,n}(\mathbf{x}) - F_\vartheta(\mathbf{x})| = o(1), \quad \forall \mathbf{x} \in \mathbb{R}^d$.

This theorem implies immediately the following corollary.

Corollary C.4. If $(P_{X^\vartheta})_{\vartheta\in\Theta}$ is uniformly absolutely continuous over Θ with respect to some continuous probability measure Q, then the following two statements are equivalent:

1. $P_{X_n^\vartheta} \overset{w,\Theta}{\Longrightarrow} P_{X^\vartheta}$;
2. $X_n^\vartheta \overset{D,\Theta}{\Longrightarrow} X^\vartheta$.

Proof of Theorem C.3. That 1. implies 2. follows immediately from 7. of Lemma C.1, since the sets

$$A_{\mathbf{x}} = (-\infty, \mathbf{x}], \quad \mathbf{x} \in \mathbb{R}^d$$

are such that $\sup_{\vartheta\in\Theta}\mu^{\vartheta}(\partial A_{\mathbf{x}})=0$, due to continuity of $(\mu^{\vartheta})_{\vartheta\in\Theta}$. Thus,

$$\sup_{\vartheta\in\Theta}\left|F_{\vartheta,n}(\mathbf{x})-F_{\vartheta}(\mathbf{x})\right|=\sup_{\vartheta\in\Theta}\left|\mu_n^{\vartheta}(A_{\mathbf{x}})-\mu^{\vartheta}(A_{\mathbf{x}})\right|=o(1).$$

To verify that 2. implies 1. we make use of Lemma C.2 by considering

$$\tilde{\mathcal{B}}=\{(\mathbf{a},\mathbf{b})\mid \mathbf{a}<\mathbf{b},\mathbf{a},\mathbf{b}\in\mathbb{R}^d\}$$

which satisfies the assumptions of Lemma C.2 if $\mathcal{A}$ is the Borel-σ-algebra on $\mathbb{R}^d$. Note that

$$\mu_n^{\vartheta}\left((\mathbf{a},\mathbf{b})\right)=F_{\vartheta,n}(\mathbf{b}-)-F_{\vartheta,n}(\mathbf{a}),\quad\text{and}\quad \mu^{\vartheta}\left((\mathbf{a},\mathbf{b})\right)=F_{\vartheta}(\mathbf{b})-F_{\vartheta}(\mathbf{a}),$$

where $F_{\vartheta,n}(\mathbf{b}-)=\lim_{\mathbf{x}\nearrow\mathbf{b}}F_{\vartheta,n}(\mathbf{x})$. Thus, it suffices to show that

$$\sup_{\vartheta\in\Theta}\left|F_{\vartheta,n}(\mathbf{x}-)-F_{\vartheta}(\mathbf{x})\right|=o(1),\quad \forall\mathbf{x}\in\mathbb{R}^d.$$

For this purpose, firstly obtain by monotonicity

$$\limsup_n\sup_{\vartheta\in\Theta}\left(F_{\vartheta,n}(\mathbf{x}-)-F_{\vartheta}(\mathbf{x})\right)\leq\limsup_n\sup_{\vartheta\in\Theta}\left(F_{\vartheta,n}(\mathbf{x})-F_{\vartheta}(\mathbf{x})\right)=0.$$

Secondly, we have

$$\liminf_n\inf_{\vartheta\in\Theta}\left(F_{\vartheta,n}(\mathbf{x}-)-F_{\vartheta}(\mathbf{x})\right)\geq 0,$$

which concludes the proof. Indeed, let $\varepsilon>0$ then there exists by the uniform absolute continuity of $(\mu^{\vartheta})_{\vartheta\in\Theta}$ over Θ w.r.t μ a $\delta>0$ such that

$$F_{\vartheta}\left(\mathbf{x}-\delta(1,\ldots,1)^T\right)\geq F_{\vartheta}(\mathbf{x})-\varepsilon,\quad\forall\vartheta\in\Theta,$$

due to continuity of μ. Thus by 2. there exists $N\in\mathbb{N}$ such that $F_{\vartheta,n}\left(\mathbf{x}-\delta(1,\ldots,1)^T\right)\geq F_{\vartheta}(\mathbf{x})-2\varepsilon$ for any $n\geq N$ and any $\vartheta\in\Theta$. By monotonicity obtain

$$\liminf_n\inf_{\vartheta\in\Theta}\left(F_{\vartheta,n}(\mathbf{x}-)-F_{\vartheta}(\mathbf{x})\right)\geq-2\varepsilon,$$

which yields the assertion by considering $\varepsilon\to 0$. □

Approximation result

The next theorem is most useful to verify uniform convergence in distribution for two sequences of random vectors if the uniform distributional limit for one sequence is known and the distance between the random vectors is uniformly tending to zero in probability.

Theorem C.5. *Let for any $\vartheta\in\Theta$, $(Y_n^{\vartheta})_{n\in\mathbb{N}}$ be a sequence of real-valued random vectors in $\mathbb{R}^d$. Suppose that $X_n^{\vartheta}\overset{D,\Theta}{\Longrightarrow}X^{\vartheta}$ and $\|Y_n^{\vartheta}-X_n^{\vartheta}\|=o_{P,\Theta}(1)$ and in addition $(P_{X^{\vartheta}})_{\vartheta\in\Theta}$ is uniformly absolutely continuous over Θ with respect to some continuous probability measure Q. Then,*

$$Y_n^{\vartheta}\overset{D,\Theta}{\Longrightarrow}X^{\vartheta}.$$

Proof of Theorem C.5. Let $g:\mathbb{R}^d\to\mathbb{R}$ be a bounded, Lipschitz-continuous function with Lipschitz constant $L > 0$. Let $\varepsilon > 0$, then

$$\sup_{\vartheta\in\Theta}\Big|\int g\,\mathrm{d}P_{Y_n^\vartheta} - \int g\,\mathrm{d}P_{X_n^\vartheta}\Big| \le L\varepsilon + 2\|g\|_\infty \sup_{\vartheta\in\Theta} P_\vartheta\big(\|X_n^\vartheta - Y_n^\vartheta\| > \varepsilon\big).$$

Thus,

$$\begin{aligned}&\sup_{\vartheta\in\Theta}\Big|\int g\,\mathrm{d}P_{Y_n^\vartheta} - \int g\,\mathrm{d}P_{X^\vartheta}\Big| \\ &\qquad \le L\varepsilon + 2\|g\|_\infty \sup_{\vartheta\in\Theta} P_\vartheta\big(\|X_n^\vartheta - Y_n^\vartheta\| > \varepsilon\big) + \sup_{\vartheta\in\Theta}\Big|\int g\,\mathrm{d}P_{X_n^\vartheta} - \int g\,\mathrm{d}P_{X^\vartheta}\Big|.\end{aligned}$$

The second term on the right-hand side of the latter inequality tends to zero, as does the third by 2. of Lemma C.1. Considering $\varepsilon \to 0$ yields that $\sup_{\vartheta\in\Theta}|\int g\,dP_{Y_n^\vartheta} - \int g\,\mathrm{d}P_{X^\vartheta}| = o(1)$. Now, Corollary C.4 and 2. of Lemma C.1 complete the proof. ◻

Uniform version of Lévy's continuity theorem

Lemma C.6. *Suppose $(P_{X^\vartheta})_{\vartheta\in\Theta}$ is uniformly absolutely continuous over Θ with respect to some continuous probability measure Q. Let Y be some random vector in $\mathbb{R}^d$. If*

$$P_{X_n^\vartheta+\sigma Y} \overset{w,\Theta}{\Longrightarrow} P_{X^\vartheta+\sigma Y}, \quad \forall\sigma > 0$$

then $P_{X_n^\vartheta} \overset{w,\Theta}{\Longrightarrow} P_{X^\vartheta}$.

Proof of Lemma C.6. Let $g : \mathbb{R}^d \to \mathbb{R}$ be a bounded and continuous function. For all $\varepsilon > 0$ there exists a $\delta > 0$ such that if $\|\mathbf{x} - \mathbf{z}\| < \delta$ then $|g(\mathbf{x}) - g(\mathbf{z})| < \varepsilon/6$. Let σ be small enough such that $\sup_{\vartheta\in\Theta} P_{X^\vartheta}\big(\|Y\| \ge \delta\sigma^{-1}\big) < \varepsilon/12\|g\|_\infty$. Such a σ exists due to the uniform absolute continuity of X_n^ϑ w.r.t. Q. Thus,

$$\begin{aligned}\sup_{\vartheta\in\Theta}\Big|\int g\,\mathrm{d}P_{X_n^\vartheta} - \int g\,\mathrm{d}P_{X_n^\vartheta+\sigma Y}\Big| &\le \sup_{\vartheta\in\Theta}\Big|\int 1_{\{\sigma\|Y\|<\delta\}}\, g\,\mathrm{d}(P_{X_n^\vartheta} - P_{X_n^\vartheta+\sigma Y})\Big| \\ &\quad + \sup_{\vartheta\in\Theta}\Big|\int 1_{\{\sigma\|Y\|\ge\delta\}}\, g\,\mathrm{d}\,(P_{X_n^\vartheta} - P_{X_n^\vartheta+\sigma Y})\Big| \\ &\le \varepsilon/6 + 2\|g\|_\infty \sup_{\vartheta\in\Theta} P_{X^\vartheta}\big(\|Y\| \ge \delta\sigma^{-1}\big) \le \varepsilon/3.\end{aligned}$$

Similarly,

$$\sup_{\vartheta\in\Theta}\Big|\int g\,\mathrm{d}P_{X^\vartheta} - \int g\,\mathrm{d}P_{X^\vartheta+\sigma Y}\Big| \le \varepsilon/3.$$

Next, by assumption there exists $N \in \mathbb{N}$ such that

$$\sup_{\vartheta\in\Theta}\Big|\int g\,\mathrm{d}P_{X_n^\vartheta+\sigma Y} - \int g\,\mathrm{d}P_{X^\vartheta+\sigma Y}\Big| \le \varepsilon/3, \quad \forall n \ge N.$$

Thus, by triangle inequality

$$\sup_{\vartheta\in\Theta}\Big|\int g\,\mathrm{d}P_{X_n^\vartheta} - \int g\,\mathrm{d}P_{X^\vartheta}\Big| \le \varepsilon, \quad \forall n \ge N,$$

which concludes the lemma. ◻

The next theorem is sufficient to derive a uniform version of Lévy's continuity theorem.

Theorem C.7. *Assume that $(P_{X^\vartheta})_{\vartheta\in\Theta}$ is uniformly absolutely continuous over Θ with respect to some continuous probability measure Q. The following two statements are equivalent:*

1. $P_{X_n^\vartheta} \overset{w,\Theta}{\Longrightarrow} P_{X^\vartheta}$;
2. $\sup_{\vartheta\in\Theta} |\phi_{X_n^\vartheta}(\mathbf{t}) - \phi_{X^\vartheta}(\mathbf{t})| = o(1), \quad \forall \mathbf{t} \in \mathbb{R}^d$.

Corollary C.4 and Theorem C.7 imply the following uniform version of Lévy's continuity theorem.

Theorem C.8 (Uniform Lévy's continuity theorem)**.** *Assume $(P_{X^\vartheta})_{\vartheta\in\Theta}$ is uniformly absolutely continuous over Θ with respect to some continuous probability measure Q. Then the following two statements are equivalent:*

1. $X_n^\vartheta \overset{D,\Theta}{\Longrightarrow} X^\vartheta$;
2. $\sup_{\vartheta\in\Theta} |\phi_{X_n^\vartheta}(\mathbf{t}) - \phi_{X^\vartheta}(\mathbf{t})| = o(1), \quad \forall \mathbf{t} \in \mathbb{R}^d$.

Proof of Theorem C.7. Suppose $P_{X_n^\vartheta} \overset{w,\Theta}{\Longrightarrow} P_{X^\vartheta}$ holds. The implication

$$\sup_{\vartheta\in\Theta} |\phi_{X_n^\vartheta}(\mathbf{t}) - \phi_{X^\vartheta}(\mathbf{t})| = o(1), \quad \forall \mathbf{t} \in \mathbb{R}^d$$

follows immediately by splitting $\exp(i\mathbf{t}^T X)$ into its real and imaginary part, which are both bounded continuous functions.

Otherwise, assume that $\sup_{\vartheta\in\Theta} |\phi_{X_n^\vartheta}(\mathbf{t}) - \phi_{X^\vartheta}(\mathbf{t})| = o(1)$, for any $\mathbf{t} \in \mathbb{R}^d$. Let $\sigma > 0$ and let $Y \sim N_d(0, I_d)$ be independent of X_n^ϑ for any $\vartheta \in \Theta$ and any $n \in \mathbb{N}$. Then, for any $g : \mathbb{R}^d \to \mathbb{R}$ which is bounded and continuous obtain with Fubini's theorem

$$\begin{aligned}\int g \,\mathrm{d}P_{X_n^\vartheta + \sigma Y} &= \int \int g(\mathbf{y}) \varphi_{d,\sigma^2 I_d}(\mathbf{y} - \mathbf{x}) \,\mathrm{d}\mathbf{y} \,\mathrm{d}P_{X_n^\vartheta}(\mathbf{x}) \\ &=: \int g(\mathbf{y}) I_n^\vartheta(\mathbf{y}) \,\mathrm{d}\mathbf{y},\end{aligned}$$

where $\varphi_{d,\sigma^2 I_d}$ is the density function of $N_d(0, \sigma^2 I_d)$ and $I_n^\vartheta(\mathbf{y}) = \int \varphi_{d,\sigma^2 I_d}(\mathbf{y} - \mathbf{x}) \,\mathrm{d}P_{X_n^\vartheta}(\mathbf{x})$. Similarly, $\int g \,\mathrm{d}P_{X^\vartheta + \sigma Y} = \int g(\mathbf{y}) I^\vartheta(\mathbf{y}) \,\mathrm{d}\mathbf{y}$ with

$$I^\vartheta(\mathbf{y}) = \int \varphi_{d,\sigma^2 I_d}(\mathbf{y} - \mathbf{x}) \,\mathrm{d}P_{X^\vartheta}(\mathbf{x}).$$

By means of the inversion formula and a substitution

$$\varphi_{d,\sigma^2 I_d}(\mathbf{y} - \mathbf{x}) = (2\pi\sigma)^{-d} \int \exp(i\sigma^{-1}\mathbf{t}^T(\mathbf{y} - \mathbf{x}) - ||\mathbf{t}||^2/2) \,\mathrm{d}\mathbf{t}.$$

Thus, by Fubini's theorem

$$\begin{aligned}I_n^\vartheta(\mathbf{y}) &= (2\pi\sigma)^{-d} \int \exp(-i\sigma^{-1}\mathbf{t}^T\mathbf{y} - ||\mathbf{t}||^2/2) \int \exp(i\sigma^{-1}\mathbf{t}^T\mathbf{x}) \,\mathrm{d}P_{X_n^\vartheta}(\mathbf{x}) \,\mathrm{d}\mathbf{t} \\ &= (2\pi\sigma)^{-d} \int \exp(-i\sigma^{-1}\mathbf{t}^T\mathbf{y} - ||\mathbf{t}||^2/2) \,\phi_{X_n^\vartheta}(\sigma^{-1}\mathbf{t}) \,\mathrm{d}\mathbf{t}\end{aligned}$$

and analogously

$$I^{\vartheta}(\mathbf{y}) = (2\pi\sigma)^{-d} \int \exp(-i\sigma^{-1}\mathbf{t}^T\mathbf{y} - ||\mathbf{t}||^2/2)\,\phi_{X^{\vartheta}}(\sigma^{-1}\mathbf{t})\,\mathrm{d}\mathbf{t}.$$

Hence,

$$\begin{aligned}
&\sup_{\vartheta\in\Theta}\Big|\int g\,\mathrm{d}P_{X_n^{\vartheta}+\sigma Y} - \int g\,\mathrm{d}P_{X^{\vartheta}+\sigma Y}\Big| \\
&\quad\le (2\pi\sigma)^{-d}||g||_{\infty}\int\int |\exp(-i\sigma^{-1}\mathbf{t}^T\mathbf{y} - ||\mathbf{t}||^2/2)| \sup_{\vartheta\in\Theta}\Big|\phi_{X_n^{\vartheta}}(\sigma^{-1}\mathbf{t}) - \phi_{X^{\vartheta}}(\sigma^{-1}\mathbf{t})\Big|\,\mathrm{d}\mathbf{t}\,\mathrm{d}\mathbf{y},
\end{aligned}$$

which tends to zero for $n \to \infty$ by dominated convergence. This implies

$$P_{X_n^{\vartheta}+\sigma Y} \overset{w,\Theta}{\Longrightarrow} P_{X^{\vartheta}+\sigma Y}, \quad \forall \sigma > 0$$

and the proof is finished by using Lemma C.6. □

Uniform Cramér-Wold Theorem

The following uniform version of the Cramér-Wold Theorem follows immediately from Theorem C.8.

Theorem C.9. *Assume $(P_{X^{\vartheta}})_{\vartheta\in\Theta}$ is uniformly absolutely continuous over Θ with respect to some continuous probability measure Q. If $\mathbf{a}^T X_n^{\vartheta} \overset{D,\Theta}{\Longrightarrow} \mathbf{a}^T X^{\vartheta}$ for any $\mathbf{a} \in \mathbb{R}^d$, then $X_n^{\vartheta} \overset{D,\Theta}{\Longrightarrow} X^{\vartheta}$.*

Uniform version of the continuous mapping theorem

Lemma C.10. *Let for any $\vartheta \in \Theta$, $(Y_n^{\vartheta})_{n\in\mathbb{N}}$ be a sequence of real-valued random vectors in $\mathbb{R}^d$ and $(\mathbf{c}^{\vartheta})_{\vartheta\in\Theta}$ be deterministic real vectors in $\mathbb{R}^d$ with $Y_n^{\vartheta} = \mathbf{c}^{\vartheta} + o_{P,\Theta}(1)$. Furthermore, suppose $(P_{X^{\vartheta}})_{\vartheta\in\Theta}$ is uniformly absolutely continuous over Θ with respect to some continuous probability measure Q, and $X_n^{\vartheta} \overset{D,\Theta}{\Longrightarrow} X^{\vartheta}$. Then, it holds that*

$$(X_n^{\vartheta}, Y_n^{\vartheta}) \overset{D,\Theta}{\Longrightarrow} (X^{\vartheta}, \mathbf{c}^{\vartheta}).$$

Proof of Lemma C.10. For any $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^d$ it holds that

$$F_{(X_n^{\vartheta},\mathbf{c}^{\vartheta})}(\mathbf{x}_1,\mathbf{x}_2) = F_{X_n^{\vartheta}}(\mathbf{x}_1)\,F_{\mathbf{c}^{\vartheta}}(\mathbf{x}_2), \quad \text{and} \quad F_{(X^{\vartheta},\mathbf{c}^{\vartheta})}(\mathbf{x}_1,\mathbf{x}_2) = F_{X^{\vartheta}}(\mathbf{x}_1)\,F_{\mathbf{c}^{\vartheta}}(\mathbf{x}_2).$$

Thus,

$$\sup_{\vartheta\in\Theta}\Big|F_{(X_n^{\vartheta},\mathbf{c}^{\vartheta})}(\mathbf{x}_1,\mathbf{x}_2) - F_{(X^{\vartheta},\mathbf{c}^{\vartheta})}(\mathbf{x}_1,\mathbf{x}_2)\Big| \le \sup_{\vartheta\in\Theta}\Big|F_{X_n^{\vartheta}}(\mathbf{x}_1) - F_{X^{\vartheta}}(\mathbf{x}_1)\Big| = o(1),$$

which shows $(X_n^{\vartheta}, \mathbf{c}^{\vartheta}) \overset{D,\Theta}{\Longrightarrow} (X^{\vartheta}, \mathbf{c}^{\vartheta})$. Note that $\sup_{\vartheta\in\Theta}||(X_n^{\vartheta}, Y_n^{\vartheta}) - (X_n^{\vartheta}, \mathbf{c}^{\vartheta})|| = o_{P,\Theta}(1)$ and with Theorem C.5 the assertion follows. □

The following theorem is a uniform version of the continuous mapping theorem.

Theorem C.11 (Uniform continuous mapping theorem)**.** *Let $H : \mathbb{R}^d \to \mathbb{R}^s$ be continuous. If $(P_{X^\vartheta})_{\vartheta\in\Theta}$ is uniformly absolutely continuous over Θ with respect to some continuous probability measure Q and $X_n^\vartheta \overset{D,\Theta}{\Longrightarrow} X^\vartheta$, then*

$$H(X_n^\vartheta) \overset{D,\Theta}{\Longrightarrow} H(X^\vartheta).$$

The proof follows directly from Corollary C.4 since for any bounded continuous function $g : \mathbb{R}^d \to \mathbb{R}$ the composition $g \circ H$ is still bounded and continuous.

Uniform version of Slutzky's Theorem

Theorem C.12 (Uniform Slutzky's Theorem)**.** *Let for any $\vartheta \in \Theta$, $(Y_n^\vartheta)_{n\in\mathbb{N}}$ be a sequence of real-valued random vectors in $\mathbb{R}^d$ and $(\mathbf{c}^\vartheta)_{\vartheta\in\Theta}$ be deterministic real vectors in $\mathbb{R}^d$ with $Y_n^\vartheta = \mathbf{c}^\vartheta + o_{P,\Theta}(1)$. Furthermore, suppose $(P_{X^\vartheta})_{\vartheta\in\Theta}$ is uniformly absolutely continuous over Θ with respect to some continuous probability measure Q, and $X_n^\vartheta \overset{D,\Theta}{\Longrightarrow} X^\vartheta$, then*

$$X_n^\vartheta + Y_n^\vartheta \overset{D,\Theta}{\Longrightarrow} X^\vartheta + \mathbf{c}^\vartheta \quad \textit{and} \quad X_n^\vartheta \cdot Y_n^\vartheta \overset{D,\Theta}{\Longrightarrow} X^\vartheta \cdot \mathbf{c}^\vartheta,$$

where the multiplication is to be understood componentwise.

Proof of Theorem C.12. The functions $(\mathbf{x},\mathbf{y}) \mapsto \mathbf{x}+\mathbf{y}$ and $(\mathbf{x},\mathbf{y}) \mapsto \mathbf{x}\cdot\mathbf{y}$ are continuous, such that applying the uniform continuous mapping theorem C.11 in combination with Lemma C.10 yields the assertion. □

Uniform Lindeberg-Feller-Theorem

For the derivation of a uniform version of the Lindeberg-Feller-Theorem we need the following auxiliary results.

Lemma C.13. *For $z_1,\dots,z_n \in \mathbb{C}$ and $w_1,\dots,w_n \in \mathbb{C}$, where $\mathbb{C}$ is the field if complex numbers, with $\sup_i |\max\{z_i,w_i\}| \le \theta$ for some $\theta > 0$. Then,*

$$\Big|\prod_{i=1}^n z_i - \prod_{i=1}^n w_i\Big| \le \theta^{n-1}\sum_{i=1}^n |z_i - w_i|.$$

For the proof see for instance Lemma 3.4.3 in Durrett (2010).

Lemma C.14. *Let $r \in \mathbb{N}$. Then for any random variable X with $\mathbb{E}|X|^{r+1} < \infty$ it holds*

$$\Big|\varphi_X(t) - \sum_{k=0}^r \mathbb{E}\frac{(itX)^k}{k!}\Big| \le \mathbb{E}\min\{|tX|^{r+1}, 2|tX|^r\}.$$

A proof of this result is given for instance in Durrett (2010), see Equation (3.3.3).

With L'Hôpital's rule obtain the following result.

Lemma C.15. *Let $\vartheta \in \Theta$ and for each n let $c_{n,j}^\vartheta$, $1 \le j \le n$ be real-values with*

1. $\sup_{\vartheta\in\Theta}\max_j |c_{n,j}^\vartheta| \to 0$;

2. $\sup_{\vartheta\in\Theta}|\sum_{j=1}^n c_{n,j}^\vartheta - \lambda|$ *for some* $\lambda\in\mathbb{R}$;

3. $\sup_{\vartheta\in\Theta}\sup_n \sum_{j=1}^n |c_{n,j}^\vartheta| < \infty$.

Then, $\sup_{\vartheta\in\Theta}|\prod_{j=1}^n(1+c_{n,j}^\vartheta)-\exp(\lambda)|\to 0$.

Theorem C.16. *For each* $n\in\mathbb{N}$ *and* $\vartheta\in\Theta$ *let* $X_{n,i}^\vartheta$, $1\le i\le n$ *be centered and independent random vectors in* $\mathbb{R}^d$. *Assume that* $(P_{X_{n,i}^\vartheta})_{\vartheta\in\Theta}$ *is uniformly absolutely continuous over* Θ *with respect to some continuous probability measure* Q. *Moreover, suppose that*

1. $\sup_{\vartheta\in\Theta}||\sum_{i=0}^n \mathbb{E}X_{n,i}^\vartheta (X_{n,i}^\vartheta)^T - \Sigma|| = o(1)$, *for some semi-positive-definite matrix* $\Sigma\in\mathbb{R}^{d\times d}$;

2. *For any* $\varepsilon>0$ *it holds that* $\limsup_n \sup_{\vartheta\in\Theta}\sum_{i=1}^n \mathbb{E}\big(||X_{n,i}^\vartheta||^2 1_{||X_{n,i}^\vartheta||>\varepsilon}\big)=0$;

Then,

$$\sup_{\vartheta\in\Theta}\big|P\big(X_{n,1}^\vartheta+\ldots+X_{n,n}^\vartheta\le \mathbf{x}\big)-\Phi_\Sigma(\mathbf{x})\big| = o(1),$$

where Φ_Σ *is the cumulative distribution function of* $N(0,\Sigma)$.

Proof. By the uniform Cramér-Wold theorem C.9 it suffices to show the univariate version of Theorem C.16. Therefore, suppose that $X_{n,i}^\vartheta$ are random variables in $\mathbb{R}$ and $\Sigma=\sigma^2$ is a positive value. We show that

$$\sup_{\vartheta\in\Theta}\Big|\prod_{i=1}^n \varphi_{X_{n,i}^\vartheta}(t)-\exp\big(-{}^{t^2\sigma^2}\!/2\big)\Big| = o(1), \quad \forall t\in\mathbb{R}. \tag{C.2}$$

By the uniform Lévy-continuity Theorem C.8 this would complete the proof.

Let $t\in\mathbb{R}$ be fixed. Set $z_{n,i}^\vartheta=\varphi_{X_{n,i}^\vartheta}(t)$ and $w_{n,i}^\vartheta=(1-{}^{t^2\mathbb{E}[(X_{n,i}^\vartheta)^2]}\!/2)$. At first, note that $|z_{n,i}^\vartheta|\le 1$ for each n,i,ϑ as well as

$$\mathbb{E}[(X_{n,i}^\vartheta)^2]\le\varepsilon^2+\mathbb{E}\big[(X_{n,i}^\vartheta)^2 1_{|X_{n,i}^\vartheta|>\varepsilon}\big],$$

for any $\varepsilon>0$, so that by the second assumption

$$\sup_{\vartheta\in\Theta}\sup_i \mathbb{E}[(X_{n,i}^\vartheta)^2]\to 0. \tag{C.3}$$

This implies for n large enough that $|w_{n,i}^\vartheta|\le 1$ for all n,i and ϑ. Hence, by Lemma C.13

$$\Big|\prod_{i=1}^n z_{n,i}^\vartheta-\prod_{i=1}^n w_{n,i}^\vartheta\Big|\le\sum_{i=1}^n|z_{n,i}^\vartheta-w_{n,i}^\vartheta|. \tag{C.4}$$

By Lemma C.14 we have for any $\varepsilon>0$

$$\begin{aligned}|z_{n,i}^\vartheta-w_{n,i}^\vartheta| &\le \mathbb{E}\min\Big(|t X_{n,i}^\vartheta|^3, 2|t X_{n,i}^\vartheta|^2\Big)\\ &\le \varepsilon t^3\,\mathbb{E}\Big(|X_{n,i}^\vartheta|^2 1_{|X_{n,i}^\vartheta|\le\varepsilon}\Big)+2t^2\,\mathbb{E}\Big(|X_{n,i}^\vartheta|^2{}_{|X_{n,i}^\vartheta|>\varepsilon}\Big).\end{aligned}$$

By the assumptions 1. and 2. of the theorem in combination with (C.4) it follows that

$$\limsup_n\sup_{\vartheta\in\Theta}\Big|\prod_{i=1}^n z_{n,i}^\vartheta-\prod_{i=1}^n w_{n,i}^\vartheta\Big|\le\limsup_n\sup_{\vartheta\in\Theta}\sum_{i=1}^n|z_{n,i}^\vartheta-w_{n,i}^\vartheta|\le\varepsilon t^3\sigma^2.$$

Letting $\varepsilon \to 0$ the right-hand side of the preceding display converges to zero. To verify (C.2) it remains to show that

$$\limsup_n \sup_{\vartheta \in \Theta} \Big| \prod_{i=1}^n w_{n,i}^{\vartheta} - \exp\big(-{}^{t^2\sigma^2}\!/_2\big) \Big| = 0.$$

But this is immediate from Lemma C.15 by setting $c_{n,i}^{\vartheta} = -{}^{t^2 \mathbb{E}[(X_{n,i}^{\vartheta})^2]}\!/_2$. Indeed, (C.3) shows 1. of Lemma C.15, while assumption 2. of this theorem implies

$$\sup_{\vartheta \in \Theta} \Big| \sum_{i=1}^n c_{n,i}^{\vartheta} - \frac{\sigma^2 t^2}{2} \Big| \to 0,$$

and assumption 2. shows the last assumption of Lemma C.15. □

Bibliography

Belov, D. I. and R. D. Armstrong (2011). Distributions of the kullback–leibler divergence with applications. *British Journal of Mathematical and Statistical Psychology 64*(2), 291–309.

Bengs, V., M. Eulert, and H. Holzmann (2018). Asymptotic confidence sets for the jump curve in bivariate regression problems. *Preprint*.

Bengs, V. and H. Holzmann (2018). Adaptive confidence intervals for kink estimation. *Preprint*.

Bickel, P. J. and M. Rosenblatt (1973). On some global measures of the deviations of density function estimates. *The Annals of Statistics 1*, 1071–1095.

Birke, M., N. Bissantz, and H. Holzmann (2010). Confidence bands for inverse regression models. *Inverse Problems 26*(11), 115020.

Bissantz, N. and H. Holzmann (2008). Statistical inference for inverse problems. *Inverse Problems 24*(3), 034009.

Bull, A. D. (2012). Honest adaptive confidence bands and self-similar functions. *Electronic Journal of Statistics 6*, 1490–1516.

Butucea, C. (2001). Exact adaptive pointwise estimation on sobolev classes of densities. *ESAIM: Probability and Statistics 5*, 1–31.

Cai, T. and M. Low (2004). An adaptation theory for nonparametric confidence intervals. *The Annals of Statistics 32*(5), 1805–1840.

Cai, T., M. Low, and Y. Xia (2013). Adaptive confidence intervals for regression functions under shape constraints. *The Annals of Statistics 41*(2), 722–750.

Cheng, M. and M. Raimondo (2008). Kernel methods for optimal change-points estimation in derivatives. *Journal of Computational and Graphical Statistics 17*(1), 56–75.

Chernozhukov, V., D. Chetverikov, and K. Kato (2014). Anti-concentration and honest, adaptive confidence bands. *The Annals of Statistics 42*(5), 1787–1818.

Cobb, G. W. (1978). The problem of the nile: conditional solution to a changepoint problem. *Biometrika 65*(2), 243–251.

Delaigle, A., P. Hall, and F. Jamshidi (2015). Confidence bands in non-parametric errors-in-variables regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 77*(1), 149–169.

Dempfle, A. and W. Stute (2002). Nonparametric estimation of a discontinuity in regression. *Statistica Neerlandica 56*(2), 233–242.

Dette, H., A. Munk, and T. Wagner (1998). Estimating the variance in nonparametric regression – what is a reasonable choice? *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 60*(4), 751–764.

Doukhan, P., G. Oppenheim, and M. Taqqu (2002). *Theory and applications of long-range dependence*. Springer Science & Business Media.

Dümbgen, L. (1991). The asymptotic behavior of some nonparametric change-point estimators. *The Annals of Statistics 19*(3), 1471–1495.

Dümbgen, L. (2003). Optimal confidence bands for shape-restricted curves. *Bernoulli 9*(3), 423–449.

Durrett, R. (2010). *Probability: theory and examples*. Cambridge university press.

Eubank, R. and P. Speckman (1993). Confidence bands in nonparametric regression. *Journal of the American Statistical Association 88*(424), 1287–1301.

Eubank, R. and P. Speckman (1994). Nonparametric estimation of functions with jump discontinuities. *Lecture Notes-Monograph Series*, 130–144.

Gach, F., R. Nickl, and V. Spokoiny (2013). Spatially adaptive density estimation by localised haar projections. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques 49*(3), 900–914.

Garlipp, T. and C. Müller (2007). Robust jump detection in regression surface. *Sankhyā: The Indian Journal of Statistics 69*(1), 55–86.

Genovese, C. and L. Wasserman (2008). Adaptive confidence bands. *The Annals of Statistics 36*(2), 875–905.

Gijbels, I., P. Hall, and A. Kneip (1999). On the estimation of jump points in smooth curves. *Annals of the Institute of Statistical Mathematics 51*(2), 231–251.

Gijbels, I., P. Hall, and A. Kneip (2004). Interval and band estimation for curves with jumps. *Journal of Applied Probability 41*(A), 65–79.

Giné, E. and R. Nickl (2009). An exponential inequality for the distribution function of the kernel density estimator, with applications to adaptive estimation. *Probability Theory and Related Fields 143*(3-4), 569–596.

Giné, E. and R. Nickl (2010). Adaptive estimation of a distribution function and its density in sup-norm loss by wavelet and spline projections. *Bernoulli 16*(4), 1137–1163.

Giné, E. and R. Nickl (2010). Confidence bands in density estimation. *The Annals of Statistics 38*(2), 1122–1170.

Giné, E. and R. Nickl (2015). *Mathematical foundations of infinite-dimensional statistical models*. Cambridge University Press.

Goldenshluger, A., A. Juditsky, A. Tsybakov, and A. Zeevi (2008a). Change-point estimation from indirect observations. 1. minimax complexity. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques 44*(5), 787–818.

Goldenshluger, A., A. Juditsky, A. Tsybakov, and A. Zeevi (2008b). Change-point estimation from indirect observations. 2. adaptation. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques 44*(5), 819–836.

Goldenshluger, A. and V. Spokoiny (2006). Recovering convex edges of an image from noisy tomographic data. *IEEE transactions on information theory 52*(4), 1322–1334.

Goldenshluger, A., A. Tsybakov, and A. Zeevi (2006). Optimal change-point estimation from indirect observations. *The Annals of Statistics 34*(1), 350–372.

Grégoire, G. and Z. Hamrouni (2002). Change point estimation by local linear smoothing. *Journal of Multivariate Analysis 83*(1), 56–83.

Hall, P. (1992). Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *The Annals of Statistics 20*(2), 675–694.

Hall, P., J. Kay, and D. Titterinton (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika 77*(3), 521–528.

Hall, P. and D. Titterington (1992). Edge-preserving and peak-preserving smoothing. *Technometrics 34*(4), 429–440.

Härdle, W. (1989). Asymptotic maximal deviation of m-smoothers. *Journal of Multivariate Analysis 29*(2), 163–179.

Härdle, W. and A. W. Bowman (1988). Bootstrapping in nonparametric regression: local adaptive smoothing and confidence bands. *Journal of the American Statistical Association 83*(401), 102–110.

Hinkley, D. V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika 57*(1), 1–17.

Hoffmann, M. and R. Nickl (2011). On adaptive inference and confidence bands. *The Annals of Statistics 39*(5), 2383–2409.

Huh, J. and K. Carriere (2002). Estimation of regression functions with a discontinuity in a derivative with local polynomial fits. *Statistics & Probability letters 56*(3), 329–343.

Ibragimov, I. A. and R. Z. Has' Minskii (1981). *Statistical estimation: asymptotic theory*, Volume 16. Springer Science & Business Media.

Kang, Y. and P. Qiu (2014). Jump detection in blurred regression surfaces. *Technometrics 56*(4), 539–550.

Khoshnevisan, D. (2002). *Multiparameter processes: an introduction to random fields*. Springer Science & Business Media.

Korostelev, A. and A. Tsybakov (1993). *Minimax Methods for Image Reconstruction*. Springer-Verlag New York, Inc.

Korostelev, A. P. (1987). On a minimax estimating of a discontinuous signal. *Teoriya Veroyatnostei i ee Primeneniya 32*(4), 796–799.

Lepskii, O. (1992). Asymptotically minimax adaptive estimation. i: Upper bounds. optimally adaptive estimates. *Theory of Probability & Its Applications 36*(4), 682–697.

Li, K.-C. (1989). Honest confidence regions for nonparametric regression. *The Annals of Statistics 17*(3), 1001–1008.

Lin, Z., D. Li, and J. Chen (2008). Change point estimators by local polynomial fits under a dependence assumption. *Journal of Multivariate Analysis 99*(10), 2339–2355.

Loader, C. (1996). Change point estimation using nonparametric regression. *The Annals of Statistics 24*(4), 1667–1678.

Lombard, F. (1988). Detecting change points by fourier analysis. *Technometrics 30*(3), 305–310.

Lounici, K. and R. Nickl (2011). Global uniform risk bounds for wavelet deconvolution estimators. *The Annals of Statistics 39*(1), 201–231.

Low, M. (1997). On nonparametric confidence intervals. *The Annals of Statistics 25*(6), 2547–2554.

Luan, Y. and Z. Xie (2001). The wavelet identification for jump points of derivative in regression model. *Statistics & Probability letters 53*(2), 167–180.

Mallik, A., M. Banerjee, and B. Sen (2013). Asymptotics for p-value based threshold estimation in regression settings. *Electronic Journal of Statistics 7*(1), 2477–2515.

Mammen, E. and W. Polonik (2013). Confidence regions for level sets. *Journal of Multivariate Analysis 122*(C), 202–214.

McDonald, J. A. and A. B. Owen (1986). Smoothing with split linear fits. *Technometrics 28*(3), 195–208.

Menéndez, P., S. Ghosh, and J. Beran (2010). On rapid change points under long memory. *Journal of Statistical Planning and Inference 140*(11), 3343–3354.

Mukherjee, P. S. and P. Qiu (2011). 3-d image denoising by local smoothing and nonparametric regression. *Technometrics 53*(2), 196–208.

Müller, H.-G. (1992). Change-points in nonparametric regression analysis. *The Annals of Statistics 20*(2), 737–761.

Müller, H.-G. and K. Song (1994). Maximin estimation of multidimensional boundaries. *Journal of Multivariate Analysis 50*(2), 265–281.

Müller, H.-G. and K.-S. Song (1997). Two-stage change-point estimators in smooth regression models. *Statistics & Probability Letters 34*(4), 323–335.

Müller, H.-G. and U. Stadtmüller (1999). Discontinuous versus smooth regression. *The Annals of Statistics 27*(1), 299–337.

Munk, A., N. Bissantz, T. Wagner, and G. Freitag (2005). On difference-based variance estimation in nonparametric regression when the covariate is high dimensional. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67*(1), 19–41.

Neumann, M. H. (1995). Automatic bandwidth choice and confidence intervals in nonparametric regression. *The Annals of Statistics 23*(6), 1937–1959.

Neumann, M. H. (1998). Strong approximation of density estimators from weakly dependent observations by density estimators from independent observations. *The Annals of Statistics 26*(5), 2014–2048.

Neumann, M. H. and J. Polzehl (1998). Simultaneous bootstrap confidence bands in nonparametric regression. *Journal of Nonparametric Statistics 9*(4), 307–333.

Park, C.-W. and W.-C. Kim (2004). Estimation of a regression function with a sharp change point using boundary wavelets. *Statistics & Probability letters 66*(4), 435–448.

Park, C.-W. and W.-C. Kim (2006). Wavelet estimation of a regression function with a sharp change point in a random design. *Journal of Statistical Planning and Inference 136*(7), 2381–2394.

Picard, D. and K. Tribouley (2000). Adaptive confidence interval for pointwise curve estimation. *The Annals of Statistics 28*(1), 298–335.

Porter, J. and P. Yu (2015). Regression discontinuity designs with unknown discontinuity points: Testing and estimation. *Journal of Econometrics 189*(1), 132–147.

Prieur, C. (2007). Change point estimation by local linear smoothing under a weak dependence condition. *Mathematical Methods of Statistics 16*(1), 25–41.

Proksch, K. (2016). On confidence bands for multivariate nonparametric regression. *Annals of the Institute of Statistical Mathematics 68*(1), 209–236.

Proksch, K., N. Bissantz, and H. Dette (2015). Confidence bands for multivariate and time dependent inverse regression models. *Bernoulli 21*(1), 144–175.

Qiao, W. and W. Polonik (2016). Theoretical analysis of nonparametric filament estimation. *The Annals of Statistics 44*(3), 1269–1297.

Qiu, P. (1997). Nonparametric estimation of jump surface. *Sankhyā: The Indian Journal of Statistics, Series A 59*(2), 268–294.

Qiu, P. (2002). A nonparametric procedure to detect jumps in regression surfaces. *Journal of Computational and Graphical Statistics 11*(4), 799–822.

Qiu, P. (2005). *Image Processing and Jump Regression Analysis*. John Wiley & Sons.

Qiu, P. and X. Li (1991). Estimation of jump regression function. *Bulletin of Informatics and Cybernetics 24*(3-4), 197–212.

Qiu, P. and B. Yandell (1997). Jump detection in regression surfaces. *Journal of Computational and Graphical Statistics 6*(3), 332–354.

Raimondo, M. (1998). Minimax estimation of sharp change points. *The Annals of Statistics 26*(4), 1379–1397.

Rio, E. (1993). Strong approximation for set-indexed partial sum processes via kmt constructions. i. *The Annals of Probability 21*(2), 759–790.

Robins, J. and A. van der Vaart (2006). Adaptive nonparametric confidence sets. *The Annals of Statistics 34*(1), 229–253.

Seijo, E. and B. Sen (2011). Change-point in stochastic design regression and the bootstrap. *The Annals of Statistics 39*(3), 1580–1607.

Smirnov, N. V. (1950). On the construction of confidence regions for the density of distribution of random variables. *Doklady Akad. 74*, 189–191.

Sonka, M., V. Hlavac, and R. Boyle (2014). *Image processing, analysis, and machine vision*. Cengage Learning.

Sun, J. and P. Qiu (2007). Jump detection in regression surfaces using both first-order and second-order derivatives. *Journal of Computational and Graphical Statistics 16*(2), 289–311.

Thistlethwaite, D. L. and D. T. Campbell (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology 51*(6), 309.

Tsybakov, A. B. (1998). Pointwise and sup-norm sharp adaptive estimation of functions on the sobolev classes. *The Annals of Statistics 26*(6), 2420–2469.

Tsybakov, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer-Verlag New York, Inc.

van der Vaart, A. W. (2000). *Asymptotic Statistics*, Volume 3. Cambridge university press.

van der Vaart, A. W. and J. Wellner (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer.

Vershynin, R. (2018). *High-dimensional probability*. Cambridge university press.

Viens, F. and A. Vizcarra (2007). Supremum concentration inequality and modulus of continuity for sub-nth chaos processes. *Journal of Functional Analysis 248*(1), 1–26.

Von Neumann, J. (1941). Distribution of the ratio of the mean square successive difference to the variance. *The Annals of Mathematical Statistics 12*(4), 367–395.

Wang, Y. (1995). Jump and sharp cusp detection by wavelets. *Biometrika 82*(2), 385–397.

Wang, Y. (1998). Change curve estimation via wavelets. *Journal of the American Statistical Association 93*(441), 163–172.

Wishart, J. (2009). Kink estimation with correlated noise. *Journal of the Korean Statistical Society 38*(2), 131–143.

Wishart, J. (2011). Minimax lower bound for kink location estimators in a nonparametric regression model with long-range dependence. *Statistics & Probability Letters 81*(12), 1871–1875.

Wishart, J. and R. Kulik (2010). Kink estimation in stochastic regression with dependent errors and predictors. *Electronic Journal of Statistics 4*, 875–913.

Wu, J. and C. Chu (1993). Kernel-type estimators of jump points and values of a regression function. *The Annals of Statistics 21*(3), 1545–1566.

Yin, Y. (1988). Detection of the number, locations and magnitudes of jumps. *Communications in Statistics. Stochastic Models 4*(3), 445–455.

Selbständigkeitserklärung

Ich erkläre, dass ich die vorliegende Dissertation

Confidence sets for change-point problems in nonparametric regression

selbständig und ohne fremde Hilfe verfasst habe. Alle verwendeten Quellen und Hilfsmittel habe ich explizit angegeben und wörtliche oder sinngemäße Zitate als solche gekennzeichnet.

Diese Dissertation wurde bislang weder in der vorliegenden noch in einer ähnlichen Form bei einer in- oder ausländischen Hochschule anlässlich eines Promotionsgesuchs oder zu anderen Prüfungszwecken eingereicht.

Ich erkläre, dass dies mein erster Versuch einer Promotion ist.

Viktor Bengs